THAÍS CAROLINA DA SILVA DAL'SASSO

EVOLUTIONARY HISTORY OF THE NECROSIS- AND ETHYLENE-INDUCING LIKE PROTEIN SUPERFAMILY ACROSS THE DOTHIDEOMYCETES (ASCOMYCOTA)

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

Orientador: Luiz Orlando de Oliveira

VIÇOSA – MINAS GERAIS
2021

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
D151e
2021

Dal'Sasso, Thaís Carolina da Silva, 1989-
Evolutionary history of the necrosis- and ethylene-inducing like protein superfamily across the Dothideomycetes (Ascomycota) / Thaís Carolina da Silva Dal'Sasso. – Viçosa, MG, 2021.
121 f. : il. (algumas color.) ; 29 cm.

Texto em inglês.
Orientador: Luiz Orlando de Oliveira.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Fungos fitopatogênicos. 2. Evolução. 3. Genômica. 4. Filogenia. 5. Genômica comparativa. 6. *Corynespora cassiicola*. I. Universidade Federal de Viçosa. Departamento de Bioquímica e Biologia Molecular. Programa de Pós-Graduação em Genética e Melhoramento. II. Título.

CDD 22. ed. 579.564

Bibliotecário(a) responsável: Bruna Silva CRB6/2552

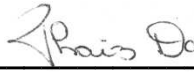
THAÍS CAROLINA DA SILVA DAL'SASSO

EVOLUTIONARY HISTORY OF THE NECROSIS- AND ETHYLENE-INDUCING LIKE PROTEIN SUPERFAMILY ACROSS THE DOTHIDEOMYCETES (ASCOMYCOTA)

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 25 de junho de 2021.

Assentimento:



Thaís Carolina da Silva Dal'Sasso
Autora

Luiz Orlando de Oliveira
Orientador

AGRADECIMENTOS

A Deus pelo dom da vida e pelas inúmeras bênçãos concedidas.

À minha mãe, Maria da Aparecida, pelo amor incondicional e por sempre lutar junto comigo para que eu alcançasse meus objetivos e realizasse os meus sonhos.

Ao professor Luiz Orlando de Oliveira pela orientação, ensinamentos, confiança, amizade e por ser um exemplo de profissionalismo.

Ao doutor Hugo Rody por aceitar ser membro da minha comissão orientadora, pelos ensinamentos e, sobretudo, pela sua amizade.

A todos os amigos do Laboratório de Biologia Molecular e Filogeografia com quem tive o prazer de conviver durante esses anos.

À Universidade Federal de Viçosa e todos os seus professores e funcionários que contribuíram para o meu crescimento profissional e pessoal.

Ao programa de Pós-Graduação em Genética e Melhoramento (PPGGM) por me permitir aprimorar meus conhecimentos e desenvolver o presente trabalho.

A todos os funcionários, estudantes e professores do Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), pela disponibilidade e ajuda sempre que necessário.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão de bolsas de estudos.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e ao CNPq por financiamentos que tornaram a realização deste trabalho possível.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

RESUMO

DAL'SASSO, Thaís Carolina da Silva, D.Sc., Universidade Federal de Viçosa, junho de 2021. **História evolutiva da superfamília de Proteínas Indutoras de Necrose e Etileno nos Dothideomycetes (Ascomycota).** Orientador: Luiz Orlando de Oliveira.

Proteínas indutoras de necrose e etileno (NLPs) formam uma superfamília amplamente distribuída e estão envolvidas nas fases iniciais da infecção de plantas, podendo desencadear respostas imunes do hospedeiro. A maioria das NLPs possui atividade citotóxica que induz a morte celular em eudicotiledôneas. Dothideomycetes constituem uma grande e diversa classe dentro do filo Ascomycota. *Corynespora cassiicola* (Pleosporales) é um fitopatógeno polífago, que causa doenças em diversas culturas. Nós utilizamos abordagens de genômica comparativa e análises de filogenia molecular para investigar a história evolutiva das NLPs nos Dothideomycetes e em *C. cassiicola*. Inicialmente, descrevemos dois genomas: *C. cassiicola* CC_29 e *C. olivacea* CBS 114450. Nós mostramos que anotações para os dois genomas diferiram e nossa filogenômica indicou que *Corynespora* não é um gênero monofilético e sugerimos que a classificação de *C. olivacea* neste gênero seja revisada. Nós investigamos a evolução das NLPs entre os Dothideomycetes. Nós acessamos os recursos genômicos de 33 espécies (cinco ordens) com diferentes estilos de vida. As filogenias desbalanceadas das NLPs sugeriram caminhos evolutivos independentes para suas subfamílias dentro dos Dothideomycetes, cada uma com diferentes assinaturas de antigas duplicações gênicas e sucessivas perdas tendenciosas de genes que podem estar associadas à atividade citotóxica. Posteriormente, acessamos os recursos genômicos de 44 isolados de *C. cassiicola* e os perfis de expressão gênica dos genes NLP de *C. cassiicola* CC_29 nas primeiras horas de colonização do hospedeiro. Três genes que codificam efetores NLP foram mantidos em *C. cassiicola* e evoluíram sob diferentes restrições seletivas. Apesar de altamente divergentes, os efetores NLP de *C. cassiicola* mantiveram conservados os resíduos necessários para desencadear as respostas imunes das plantas. No entanto, os perfis de expressão gênica indicaram que apenas um gene efetor NLP é altamente expresso nas primeiras horas de colonização de soja por *C. cassiicola* CC_29.

Palavras-chave: Efetores. Evolução. Filogenômica. Genômica comparativa. *Corynespora cassiicola.*

ABSTRACT

DAL'SASSO, Thaís Carolina da Silva, D.Sc., Universidade Federal de Viçosa, June 2021. **Evolutionary history of the Necrosis- and Ethylene-inducing Like Protein superfamily across the Dothideomycetes (Ascomycota**). Adviser: Luiz Orlando de Oliveira.

Necrosis- and Ethylene-inducing peptide 1-like proteins (NLPs) form a broadly distributed superfamily involded in the early phases of plant infection and may trigger host immune responses. The majority of NLPs acts as cytotoxic proteins, inducing cell-death in eudicot plants. Dothideomycetes is a large and diverse class within the phylum Ascomycota. *Corynespora cassiicola* (Pleosporales) is a widespread, polyphagous plant-pathogen that causes leaf diseases on many crops. We used comparative genomic approaches and molecular phylogenetic analyses to investigate the evolutionary history of NLP across the Dothideomycetes and within *C. cassiicola*. Initially, we described two genomes: *C. cassiicola* CC_29 and *C. olivacea* CBS 114450. We showed that the two genome annotations differed greatly. A robust phylogenomics indicated the non-monophyly for the genus *Corynespora* and we suggested that the classification of *C. olivacea* into that genus should be revised. Subsequently, we investigate the evolution of NLPs across the Dothideomycetes. We accessed genomic resources from 33 species (five orders) with different lifestyles. The imbalanced phylogenies of NLPs suggested NLP subfamilies underwent independent evolutionary paths in the Dothideomycetes, each of which showing different signatures of ancient gene duplications and biased successive gene losses that may be associated with changes in cytotoxic activity. Subsequently, we accessed genomic resources from 44 isolates of *C. cassiicola* and the expression profiles of NLP-encoding genes of *C. cassiicola* CC_29 at the early hours of host colonization. Three NLP effector-encoding genes were maintained in the genomes of *C. cassiicola* and have evolved under different selective constraints. Despite highly divergent, NLP effectors maintained conserved the residues necessary to trigger plant immune responses. Nevertheless, gene expression profiles suggested that only one NLP effector gene is highly expressed at the early hours of soybean colonization by *C. cassiicola.*

Keywords: Effector. Evolution. Phylogenomics. Comparative genomics. *Corynespora cassiicola.*

LISTA DE SIGLAS E ABREVIATURAS

AAs	Auxiliary Activities
ANOVA	Analysis of Variance
BEB	Bayes Empirical Bayes
BIC	Bayesian Information Criteria
BUSCOs	Benchmarking Universal Single-Copy Orthologs
CBM	Carbohydrate-Binding Modules
CDS	Coding-DNA Sequence
CEs	Carbohydrate Esterases
ESS	Effective Sample Size
GHs	Glycoside Hydrolases
GTs	Glycosyl Transferases
HGT	Horizontal Gene Transfer
HPI	Hours Post Inoculation
KOG	euKaryotic Orthologous Group
LTRs	Likelihood Ratio Tests
MAMPs	Microbe-Associated Molecular Patters
MKT	McDonald and Kreitman Test
MYA	Million Years Ago
NI	Neutrality Index
NLP	Necrosis- and Ethylene-Inducing Peptide 1-Like Protein
NRPS	Non-Ribosomal Peptide Synthases
PCR	Polymerase Chain Reaction
PKS	Polyketide Synthases
PLs	Polysaccharide Lyases
PP	Posterior Probability
SCO	Single Copy Ortholog

SUMÁRIO

GENERAL INTRODUCTION

GENERAL INTRODUCTION

Plant pathogenic organisms, such as fungi, have several strategies to infect the host. Regardless of the adaptive strategy of any pathogen, they encode effector proteins in order to suppress the immune responses of the host and allow the pathogen to support growth in host tissue (Lo Presti et al., 2015; Wirthmueller et al., 2013). In response, throughout the plant-pathogen coevolutionary process, plants have developed defense mechanisms that restrict the proliferation of pathogens in their tissues (Chisholm et al., 2006; Jones and Dangl, 2006). In these plant-pathogen interactions, the emergence of new effector genes occurs as a continuous process, resulting from the strong selection pressure to which those genes are submitted throughout the evolutionary process (Franceschetti et al., 2017; Sánchez-Vallet et al., 2018).

Adaptation to new hosts and evasion of recognition by the pathogen may require rapid changes in the effector repertoire, leading to the rapid evolution of these genes (Lo Presti et al., 2015; Möller and Stukenbrock, 2017; Rovenich et al., 2014; Sánchez-Vallet et al., 2018). Evolution towards evasion of recognition and function optimization can occur by modifying nucleotide sequences, deleting genes and altering the expression profiles of existing effectors (Fouché et al., 2018; Sánchez-Vallet et al., 2018). Additionally, long-term pathogenicity may depend on the continued emergence of new effectors to be able to replace an effector that has been lost or to capture new hosts (Sánchez-Vallet et al., 2018).

The acquisition of effector genes can occur through different mechanisms. The acquisition of effectors by horizontal gene transfer (HGT) usually results in specific effectors of a few phylogenetic lineages (Fouché et al., 2018; Sánchez-Vallet et al., 2018). Additionally, new effectors can also be acquired through gene duplications, followed sequence divergence, and rearrangements in certain regions of the genome, which can lead to the gain of a sequence that encodes a signal peptide (Fouché et al., 2018; Möller and Stukenbrock, 2017; Plissonneau et al., 2017). The rapid evolution of effector genes usually results in the predominance of small-sized amino acid sequences and the absence of conserved domain effector proteins (Jones et al., 2018; Selin et al., 2016; Sonah et al., 2016; Sperschneider et al., 2015). For this reason, the prediction of fungal effectors is considered a highly complex task.

The prediction of effector genes is done through the use of computational tools and bioinformatic analyses that consider properties that are common to most effectors (Sperschneider et al., 2015). Generally, sets of candidate effectors are identified from a genome, iteratively filtering a predicted proteome for secreted proteins and considering the following general properties: a) presence of a signal peptide for secretion; b) absence of transmembrane domains; c) small size (<300 amino acids) or low molecular weight of the predicted protein (<30kDa); d) cysteine richness (usually above four residues); e) absence of PFAM domains, except in cases when they are associated with pathogenicity; f) genes belonging to duplicated families (Jones et al., 2018; Selin et al., 2016; Sonah et al., 2016; Sperschneider et al., 2015). The validation of effectors that meet these properties can be conducted through comparative genomic analyses or *in planta* expression data (Jones et al., 2018; Sperschneider et al., 2015).

The difficulty in recognizing common domains that indicate effector function or activity is due to their enzymatic activity or the absence of domains known for direct interaction with host factors (Rovenich et al., 2014). Nevertheless, some proteins secreted by pathogens may contain domains that target specific host proteins (Franceschetti et al., 2017; Liu et al., 2019). The discovery of effector proteins with conserved domains initiated a new direction in plant-pathogen interaction studies, since the identification of these domains in the amino acid sequence makes it possible to determine the structure and, consequently, predict the function of these proteins in the host cell (Jones et al., 2018; Sonah et al., 2016).

The advances in effector identification and validation studies in recent years are due to a combination of factors. Advances in sequencing technologies have allowed the generation of large volumes of data at increasingly reduced time and costs (Jones et al., 2018). Additionally, the development of increasingly efficient and robust computational tools for prediction of effector genes has allowed accurate analyses of the big datasets (Sperschneider et al., 2015). The application of comparative genomics analyses at inter and intraspecific levels has contributed to the identification of new effectors and to our understanding about the acquisition and evolution processes of these genes in the genomes of plant pathogens (Plissonneau et al., 2017; Snelders et al., 2018). Therefore, expanding the knowledge of how the evolution of effectors are linked to the evolution of the pathogen represents an important area for future researches understand the physiological processes of disease development in plants.

THESIS STRUCTURE AND GENERAL OBJECTIVES

This thesis consists of three chapters.

In the first chapter, we described two genome sequences and detailed their effector repertoires. We provided genomic data for *Corynespora cassiicola* isolate CC_29 (isolated from soybeans) and the first genome of *Corynespora olivacea* isolate CBS 114450. We used several computational tools in order to performed functional annotations and unravel the effectorome of both genomes. Additionally, we used a phylogenomic approach to investigate the phylogenetic relationships between the two congeners in an attempt to shed light into the current classification of *C. olivacea* into the genus *Corynespora*.

In the second chapter, we accessed the genomic resources of 33 species from five orders in the Dothideomycetes class of fungi to study the evolutionary history of the Necrosis- and Ethylene-inducing peptide 1-like protein (NLP) superfamily. Using phylogenetic frameworks, we investigate whether biased gene retention have droved the NLP evolution. Additionally, we explore whether the cytotoxic activities reported through functional analyses are associated with NLP gene evolution and phylogenetic relationships among species of Dothideomycetes.

In the third chapter, we investigated the evolutionary history of NLPs in a species-limited framework. We accessed the genomic resources of 44 isolates of *C. cassiicola* to infer the gene genealogy of NLP-encoding genes and determined the evolutionary forces that shaped gene evolution of NLP effectors. We also evaluated the expression profiles of NLP effector genes by *C. cassiicola* during the early phase of host colonization.

REFERENCES

Chisholm, S.T., Coaker, G., Day, B., Staskawicz, B.J., 2006. Host-microbe interactions: Shaping the evolution of the plant immune response. Cell 124, 803–814.

Fouché, S., Plissonneau, C., Croll, D., 2018. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. Curr. Opin. Microbiol. 46, 34–42.

Franceschetti, M., Maqbool, A., Jiménez-Dalmaroni, M.J., Pennington, H.G., Kamoun, S., Banfield, M.J., 2017. Effectors of filamentous plant pathogens: Commonalities amid diversity. Microbiol. Mol. Biol. Rev. 81, 1–17.

Jones, D.A., Bertazzoni, S., Turo, C.J., Syme, R.A., Hane, J.K., 2018. Bioinformatic prediction of plant–pathogenicity effector proteins of fungi. Curr. Opin. Microbiol. 46, 43–49.

Jones, J.D., Dangl, J., 2006. The plant immune system. Nature 444, 323–329.

Liu, L., Xu, L., Jia, Q., Pan, R., Oelmüller, R., Zhang, W., Wu, C., 2019. Arms race: Diverse effector proteins with conserved motifs. Plant Signal. Behav. 14.

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., Kahmann, R., 2015. Fungal effectors and plant susceptibility. Annu. Rev. Plant Biol. 66, 513–545.

Möller, M., Stukenbrock, E.H., 2017. Evolution and genome architecture in fungal plant pathogens. Nat. Rev. Microbiol. 15, 756–771.

Plissonneau, C., Benevenuto, J., Mohd-Assaad, N., Fouché, S., Hartmann, F.E., Croll, D., 2017. Using population and comparative genomics to understand the genetic basis of effector-driven fungal pathogen evolution. Front. Plant Sci. 8, 1–15.

Rovenich, H., Boshoven, J.C., Thomma, B.P.H.J., 2014. Filamentous pathogen effector functions: Of pathogens, hosts and microbiomes. Curr. Opin. Plant Biol. 20, 96–103.

Sánchez-Vallet, A., Fouché, S., Fudal, I., Hartmann, F.E., Soyer, J.L., Tellier, A., Croll, D., 2018. The genome biology of effector gene evolution in filamentous plant pathogens. Annu. Rev. Phytopathol. 56.

Selin, C., de Kievit, T.R., Belmonte, M.F., Fernando, W.G.D., 2016. Elucidating the role of effectors in plant-fungal interactions: Progress and challenges. Front. Microbiol. 7, 1–21.

Snelders, N.C., Kettles, G.J., Rudd, J.J., Thomma, B.P.H.J., 2018. Plant pathogen effector proteins as manipulators of host microbiomes? Mol. Plant Pathol. 19,

Sonah, H., Deshmukh, R.K., Bélanger, R.R., 2016. Computational prediction of effector proteins in fungi: Opportunities and challenges. Front. Plant Sci. 7, 1–14.

Sperschneider, J., Dodds, P.N., Gardiner, D.M., Manners, J.M., Singh, K.B., Taylor, J.M., 2015. Advances and challenges in computational prediction of effectors from plant pathogenic fungi. PLoS Pathog. 11, 1–7.

Wirthmueller, L., Maqbool, A., Banfield, M.J., 2013. On the front line: Structural insights into plant-pathogen interactions. Nat. Rev. Microbiol. 11, 761–776.

Chapter 1

Genome sequences and *in silico* effector mining of *Corynespora cassiicola* CC_29 and *Corynespora olivacea* CBS 114450

Genome sequences and *in silico* effector mining of *Corynespora cassiicola* CC_29 and *Corynespora olivacea* CBS 114450

Thaís Carolina da Silva Dal'Sasso[1], Hugo Vianna Silva Rody[2], Pablo Enrique Grijalba[3], Luiz Orlando de Oliveira[1]

[1]Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Viçosa, Viçosa, Brazil
[2]Departamento de Genética, Universidade de São Paulo/Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, Brazil
[3]Facultad de Agronomía, Universidad de Buenos Aires, Buenos Aires, Argentina

Corresponding author: Luiz Orlando de Oliveira
Email: lorlando@ufv.br

ABSTRACT

The placement of *Corynespora olivacea* within the large genus *Corynespora* (Pleosporales) is controversial because the species is distantly related to other congeners, including the type species *C. cassiicola*. *Corynespora cassiicola* is a polyphagous, cosmopolitan plant pathogen. Successful colonization of plant tissues requires the pathogen's effector repertoire to modulate host cell physiology and facilitate the infection process. We sequenced and performed functional annotations on the genomes of *C. cassiicola* CC_29 (genome size about 44.8 Mb; isolated from soybean leaves) and *C. olivacea* CBS 114450 (32.3 Mb). Our phylogenomic approach showed that *C. cassiicola* is distantly related to *C. olivacea*, which clustered among the *Massarinaceae* family members, supporting a hypothesis that *C. olivacea* was originally misclassified. The predicted sizes for the proteome and secretome of *C. cassiicola* (18,487 and 1,327; respectively) were larger than those of *C. olivacea* (13,501 and 920; respectively). *Corynespora cassiicola* had a richer repertoire of effector proteins (CAZymes, proteases, lipases, and effectors) and genes associated with secondary metabolism than did *C. olivacea*.

Keywords: Genome assembly, protein annotation, secondary metabolism, phylogenomics, plant-pathogen interaction

INTRODUCTION

The genus *Corynespora* belongs to the *Corynesporascaceae* family, in the order Pleosporales (Dothideomycetes, Ascomycota) (Voglmayr and Jaklitsch, 2017), and is comprised of more than one hundred fungal species, most of which are important plant pathogens (Dixon et al., 2009). *Corynespora cassiicola* (Berk. & M.A. Curtis) C.T. Wei is the type species of the genus *Corynespora* (Schoch et al., 2009; Voglmayr and Jaklitsch, 2017), has a cosmopolitan occurrence, and is able to infect leaves, stems, and roots of more than 530 species within 380 genera of plants (Dixon et al., 2009). Rare cases of human infections have been reported (Looi et al., 2017). As a plant pathogen, *C. cassiicola* is responsible for emerging leaf diseases in several crops, such as rubber tree, soybean, cotton, tomato, cucumber, and papaya (Dixon et al., 2009; Lopez et al., 2018; Sumabat et al., 2018). In both Brazil and the United States, *C. cassiicola* has been responsible for increased yield losses of soybeans and cotton, ranking as a pathogen of major economic relevance (Molina et al., 2019; Sumabat et al., 2018).

Corynespora olivacea (Wallr.) M.B. Ellis occurs on twigs of *Tilia* spp. (*Malvaceae*) and is distributed across the temperate regions of Europe and North America (Crous et al., 2011; Voglmayr and Jaklitsch, 2017). Little is known about the genetics of *C. olivacea*. Although the species remains within the genus *Corynespora*, phylogenetic analyses based on a few nuclear genes have shown that it is actually distantly related to other congeners, such as *C. cassiicola* (Schoch et al., 2009; Tanaka et al., 2015). Intriguingly, *C. olivacea* was found phylogenetically more closely related to species within the *Massarinaceae* family (Schoch et al., 2009; Tanaka et al., 2015; Voglmayr and Jaklitsch, 2017). Such findings suggested that the current systematic placement of *C. olivacea* needs to be revised (Tanaka et al., 2015; Voglmayr and Jaklitsch, 2017).

The successful colonization of plant tissues depends upon the ability of the pathogen to either escape or suppress the several lines of host defenses (Lo Presti et al., 2015). Effectors are broadly defined as proteins capable of modulating host cell physiology or local microbial communities to promote host susceptibility (Snelders et al., 2018). Therefore, the use of computational tools for effector mining has been widely applied as a first step prior to experimental validations and development of crop resistance strategies (Sonah et al., 2016; Sperschneider et al., 2018). Although

genomic data from *C. cassiicola* have been published from isolates obtained from some host species (Gao et al., 2020; Looi et al., 2017; Lopez et al., 2018), only two draft genomes are currently available from soybean isolates from Brazil (Lopez et al., 2018). To date, no genome sequences have been published for *C. olivacea*. The availability of additional genomes of *C. cassiicola* isolates, alongside the genomes from other *Corynespora* species is critical to accelerate our understanding of their economically important pathosystems. Further, whole-genome sequence data and phylogenomics have improved the resolution of phylogenetic trees used to resolve taxonomic uncertainties and support species reclassifications (Haridas et al., 2020).

Herein, we provide genomic data from two *Corynespora* species: *C. cassiicola* isolate CC_29 (isolated from soybean leaves from Brazil) and *C. olivacea* isolate CBS 114450 (acquired from a fungal repository). We also performed a series of functional annotations in both genomes to investigate their phylogenetic relationships and compared their effector repertoires. The two species differed greatly in genome sizes and effector repertoire, with *C. cassiicola* having a larger number of effectors (such as CAZymes, lipases, and proteases) capable of interacting with metabolites and enzymes of plants. Phylogenomics showed *Corynespora* as a non-monophyletic genus and suggested that the classification of *C. olivacea* in *Corynespora* should be reconsidered. Its placement within the family *Massarinaceae* seems a likely outcome.

MATERIAL AND METHODS

Fungal cultures and genomic DNA isolation

The *Corynespora cassiicola* isolate CC_29 was obtained from symptomatic leaves of soybeans. The leaves were sampled from a commercial soybean field in the state of Mato Grosso, Brazil. Pure cultures of isolate CC_29 used in this study are maintained in our laboratory. We acquired the *C. olivacea* isolate CBS 114550 from the Westerdijk Fungal Biodiversity Institute (The Netherlands). Fungal cultures were grown on PDA medium (Potato Dextrose Agar) for 10 days, at 25 ± 0.5 °C (photoperiod 12 h/12 h). Extraction of total genomic DNA from mycelia (about 200 mg from each isolate) was carried out using the CTAB method (Doyle and Doyle, 1990). The concentration of nucleic acids was determined on a NanoDrop 2000 spectrophotometer (NanoDrop Technologies).

Genome sequencing and *de novo* assembly

Whole genome sequencing was performed by Macrogen (Macrogen Inc., South Korea) using paired-end sequencing (2 x 150 bp) on an Illumina Novaseq 6000 platform (Illumina, California, USA). Libraries were prepared using the TruSeq Nano DNA Kit according to the manufacturer's preparation guide. After downloading sequence data from the service provider, we evaluated the quality of raw reads using FastQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and ran Trimmomatic v0.36 (Bolger et al., 2014) using the settings "ILLUMINACLIP:TruSeq3.fa:2:30:10", "LEADING:3", "TRAILING:3", and "MINLEN:50" to remove adaptors and low quality (Q < 30) reads. For *C. olivacea*, we observed a bimodal distribution of GC content. To check for contaminant reads, we performed a taxonomic classification of the reads using Kraken2 (Wood et al., 2019) against the minikraken2 database v2_201904. Reads classified as *Bacillus* spp. (2.21%) were considered to be contaminants and were removed from the dataset. After removal of contaminant reads, FastQC analysis indicated unimodal GC content distribution for *C. olivacea*.

De novo assembly was carried out using SPAdes v3.9.0 (Bankevich et al., 2012) for Illumina reads, with the parameters for *C. cassiicola* set as "-k 21,33,45,57,69,81,93,105,117" and the parameters for *C. olivacea* set as "-k 21,27,33,39,45,51,57,63". Additionally, we set the parameter "--careful" for both species. We accessed the quality of genome assemblies using QUAST v5.0.2 (Mikheenko et al., 2018), with default parameters for eukaryotes, and read mapping statistics using SAMtools v1.6 (Li and Durbin, 2009). Finally, we assessed the completeness of core fungal orthologs using BUSCO v4.0.5 (Simão et al., 2015) based on the dataset fungi_odb10 with 758 core orthologs.

Gene prediction and annotation

Gene annotation was performed using Augustus v3.2.2 (Stanke and Morgenstern, 2005). For training Augustus, we used pre-existing gene structures available for two genomes: *C. cassiicola* isolate CCP (GenBank assembly GCA_003016335) (Lopez et al., 2018) and *Byssothecium circinans* isolate CBS 675.92 (GenBank assembly GCA_010015675) (Haridas et al., 2020).

For gene predictions of both *C. cassiicola* CC_29 and *C. olivacea* CBS 114450, we ran Augustus using *C. cassiicola* CCP as 'reference species' with parameters "--codingseq", "--protein", and "--cds". Subsequently, we ran Augustus with the same parameters and *B. circinans* CBS 675.92 as 'reference species' for gene prediction of *C. olivacea*. Finally, we selected the gene prediction of *C. olivacea* with the highest number of predictions for its proteome annotation.

We annotated the two predicted proteomes using PfamScan with Pfam v32.0 (El-Gebali et al., 2019) and InterProScan v5.30.69 (Jones et al., 2014) with the following parameters: SMART-7.1, SUPERFAMILY-1.75, ProDom-2006.1, CDD-3.16, TIGRFAM-15.0, Pfam v31.0, Coils-2.2.1, and Gene3D-4.2.0. Distribution of euKaryotic Orthologous Group (KOG) terms were assigned to the protein sets using the eggNOG-mapper (Huerta-Cepas et al., 2017) online tool against the eggNOG 5.0 database (Huerta-Cepas et al., 2019). The presence of a signal peptide predicted by SignalP v4.1 (Petersen et al., 2011) and the absence of transmembrane domains according to TMHMM v2.0 (Krogh et al., 2001) identified secreted proteins.

We used five different methods to annotate putative effectors in both broad and narrow senses. Carbohydrate-active enzymes (CAZymes) were predicted from the two fungal proteomes using the dbCAN2 HMM profile database v7.0 (Zhang et al., 2018) and *hmmscan* from HMMER v3.1b2 (http://hmmer.org). Lipases were predicted using HMM profiles from the Lipase Engineering Database v3.0 (Fischer and Pleiss, 2003) and *hmmscan*. We used the hmmscan-parser.py script provided by dbCAN to select significant matches for CAZymes and lipases with an E-value inclusion threshold set at 1e-4. Proteases were predicted using the fungal proteomes as queries and the MEROPS database v12.1 (Rawlings et al., 2012) with BLASTp run locally (E-value set at 1e-4 as cutoff). The secondary metabolite biosynthesis gene clusters were predicted by the standalone antiSMASH v5.1.1 (Blin et al., 2019) for fungi. Effector repertoire in a narrow sense was predicted using machine learning as implemented in EffectorP v2.0 (Sperschneider et al., 2018) from the predicted secretome.

We performed the *de novo* and homology-based identification of transposable elements (TEs) and repeats in the genomes using the RepeatModeler v1.0.11 (Flynn et al., 2020) pipeline. We combined each *de novo* library to the known repeat libraries from RepBase release 20181026 and RepeatMasker.lib and used them as input for RepeatMasker v4.1.1 (http://repeatmasker.org/).

Orthogroups e phylogenomic analysis

From MycoCosm (Grigoriev et al., 2013), we retrieved protein sequences from a total of 18 species. There were 17 species from ten families within the order Pleosporales and one species (*Hysterium pulicare*) of Hysteriales that was used as outgroup (Table S1). We used OrthoFinder v2.3.3 (Emms and Kelly, 2015) to establish orthologous relationships among the proteomes of *C. olivacea* CBS 114450, *C. cassiicola* CC_29, and the 18 species set. OrthoFinder calculated length and phylogenetic distance-normalized scores from an all-versus-all alignment using DIAMOND v0.9.24 (Buchfink et al., 2014) and assumed ortholog groups (orthogroups) by Markov clustering analysis.

We used a Maximum-Likelihood phylogeny to uncover the phylogenetic relationships for the set of species described above. The single copy ortholog (SCO) proteins identified by OrthoFinder were aligned using the L-INS-I method as implemented in MAFFT v7.453 (Katoh and Standley, 2013). trimAL v1.4.rev22 (Capella-Gutiérrez et al., 2009) was used to trim alignments, with the parameters “-gt 0.7”, “-cons 80”, and “-w 5”. After trimming and concatenation, we subjected the dataset as input to IQ-TREE v1.6.11 (Nguyen et al., 2015) to estimate both the best evolutionary model and the phylogenomic tree. According to the Bayesian information criteria (BIC), IQ-TREE standard model selection indicated JTT+F+I+G4 as the best fit amino acid model. We produced a Maximum-Likelihood phylogenomic tree using 1,000 ultrafast bootstrap replicates, ten independent runs, and *Hysterium pulicare* set as outgroup. Tree was visualized in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

RESULTS AND DISCUSSION

De novo genome assembly

We assembled a genome of *C. cassiicola* (isolate CC_29, from soybean leaves) and the first genome of *C. olivacea* (isolate CBS 114450) using Illumina reads. The final genome assemblies of *C. cassiicola* CC_29 and *C. olivacea* CBS 114450 added up to 44.79 Mbp (arranged in 2,239 scaffolds; ≥ 500 bp) and 32.32 Mbp (1,490 scaffolds; ≥ 500bp), respectively (Table 1). The assembled size of the genome of *C. cassiicola* CC_29 (44.79 Mbp) was very close to the reference genome *C. cassiicola* CCP (44.85 Mb) (Lopez et al., 2018). Overall, our genome had quality that was either comparable or higher than the quality of previously available *C. cassiicola* genomes (Gao et al., 2020; Looi et al., 2017; Lopez et al., 2018). The degree of continuity was higher in *C. cassiicola* CC_29 (N_{50} 615,474 bp) than *C. olivacea* CBS 114450 (N_{50} 109,441 bp).

The BUSCO analysis identified 99.6% of the orthologs as complete for *C. cassiicola* CC_29, and 98.7% as complete for *C. olivacea* CBS 114450. The BUSCO scores indicated a high level of completeness, making the genomes suitable for further functional analyses. In the genome of *C. cassiicola* CC_29, our analyses predicted a total of 18,487 protein-encoding genes, 1,320 (7.69%) more genes than predicted for the reference genome (17,167 genes) (*C. cassiicola* CCP; Lopez et al., 2018).

We used gene models of *C. cassiicola* CCP (*Corynesporascaceae*) and *B. circinans* CBS 675.32 (*Massarinaceae*) to train Augustus and predict genes for *C. olivacea* CBS 114450. In the *C. olivacea* CBS 114450 genome, the numbers of predicted protein-encoding genes were 13,501 (using *B. circinans* CBS 675.32) and 12,326 (using *C. cassiicola* CCP). Those values suggested that *C. olivacea* CBS 114450 shared more homologs with *B. circinans* CBS 675.32 than with the congener *C. cassiicola.* Thus, our downstream protein annotations and phylogenomic analysis of *C. olivacea* CBS 114450 used the gene prediction based on *B. circinans* CBS 675.32.

Table 1. Summary of genome sequencing, assembly and annotation statistics of *C. cassiicola* CC_29 and *C. olivacea* CBS 114450.

Genomic statistics	***C. cassiicola* CC_29**	***C. olivacea* CBS 114450**
Assembly		
Total assembly size (Mbp)	44.79	32.32
# of scaffolds (>= 500 bp)	2,239	1,490
# of scaffolds (>= 10,000 bp)	205	386
# of scaffolds (>= 50,000 bp)	97	223
Largest scaffold (bp)	2,595,751	598,525
GC (%)	51.64	53.01
N_{50} (bp)	615,474	109,441
Total reads	45,167,709	25,779,927
Mapped reads (%)	99.51	99.51
Average coverage depth (X)	150	119
Gene completeness		
Complete BUSCOs (%)	99.6	98.7
Complete and single-copy BUSCOs (%)	99.6	98.4
Annotation		
# of protein-coding genes	18,487	13,501
# of PFAM annotated proteins	11,763	7,973
# of InterPro annotated proteins	13,738	9,330
# of predicted secreted proteins	1,327	920
Repeat sequences		
Total of bases masked (Mbp)	2.62	2.54
Interspersed repeats/TEs (Mbp)	2.04	2.08
Simple repeats (Mbp)	0.50	0.40
Low complexity (Mbp)	0.08	0.06
LINEs (Mbp)	0.01	0.06
LTR elements (Mbp)	1.15	0.31
DNA transposons (Mbp)	0.06	0.20

Phylogenomics

OrthoFinder uncovered 3,583 SCO proteins (a concatenated set of 2,084,531 amino acids). In the phylogenomic tree constructed from this dataset, *C. olivacea* was distantly related to *C. cassiicola*. This finding implies that *Corynespora*, as currently circumscribed, cannot be considered a monophyletic genus (Figure 1). As anticipated, the two isolates of *C. cassiicola* (CC_29 and CCP) clustered together to form the subclade *Corynesporascaceae*. Together with *B. circinans*, *C. olivacea* clustered with *Massarina eburnea* to form a subclade for the *Massarinaceae* family. The placement of *C. olivacea* within the *Massarinaceae* agreed with the previous phylogenetic studies based on sets of nuclear genes (Schoch et al., 2009, for Dothideomycetes; Tanaka et al., 2015, for suborder Massarineae).

We suggest two plausible scenarios to explain the distant phylogenetic relationship between *C. cassiicola* and *C. olivacea*: (1) misidentification of the strain CBS 114450 as *C. olivacea*, or (2) misclassification of *C. olivacea* into the genus *Corynespora*. The second scenario seems to be more realistic because the reclassification of some species of *Corynespora*, including *C. olivacea*, has been proposed recently (Tanaka et al., 2015). Moreover, both molecular and morphological data obtained from other isolates of *C. olivacea* suggested that the species should be reclassified as *Helminthosporium oligosporum* (Voglmayr and Jaklitsch, 2017). The genome sequence we presented here provides a new source of information to support the need to investigate further the classification of *C. olivacea*.

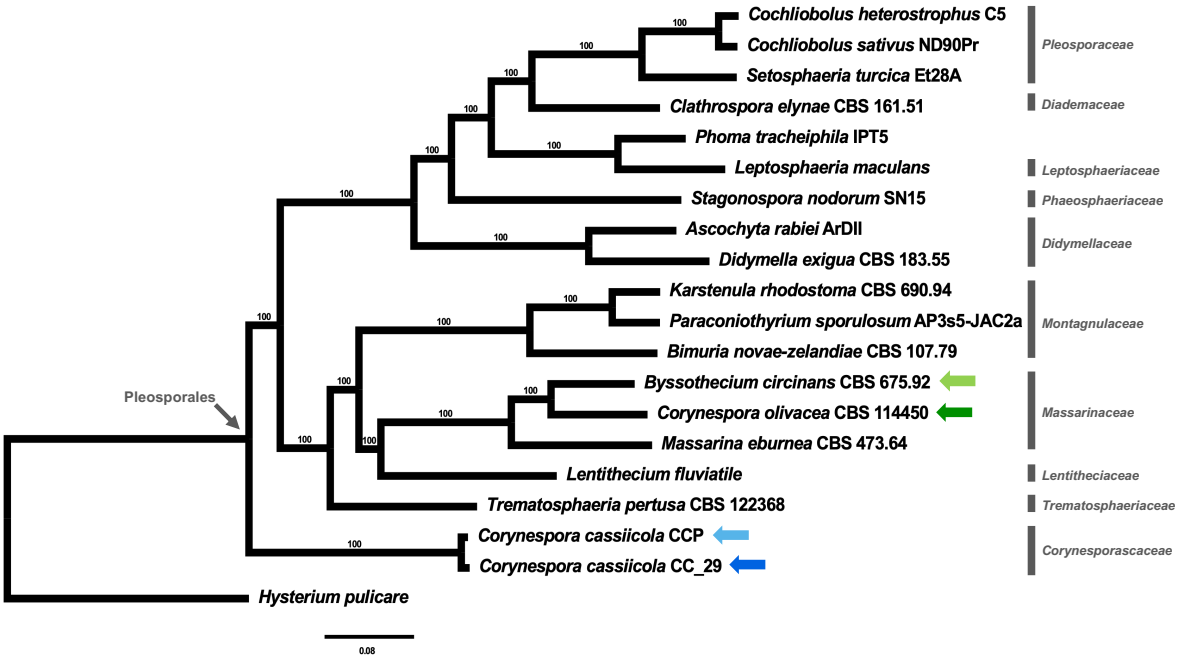


Figure 1. Phylogenomic tree (consensus tree) showing the relationships among 18 species in the order Pleosporales, with *Hysterium pulicare* (Hysteriales) used as outgroup. The maximum likelihood phylogenetic tree was based on a concatenated dataset of 3,583 single copy ortholog proteins. Arrows (and their colors) indicate specific positions in the phylogeny: blue, *Corynespora cassiicola* CC_29; light blue, *C. cassiicola* CCP; green, *C. olivacea* CBS 114450; light green, *Byssothecium circinans* CBS 675.92. Vertical bars indicate the families, as shown. Branch lengths are drawn to scale; nodal support values are given as local bootstraps above the branches. Scale bar corresponds to the expected number of substitutions per site.

Comparative and functional genomics

Compared to the smaller genome size of *C. olivacea* CBS 114450 (32.32 Mbp), the relative larger genome size of the *C. cassiicola* CC_29 (44.79 Mbp) provided more protein-encoding genes and, consequently, higher number of annotated genes/proteins (Table 1, Figure 2). In the *C. cassiicola* CC_29 genome, PFAM annotated at least one domain in 11,763 proteins (64% of a total of 18,487 proteins), whereas InterPro annotated at least one domain in 13,738 proteins (74%). Out of a total of 13,501 proteins from the *C. olivacea* CBS 114450 genome, the number of proteins that had at least one domain annotated was 7,973 (59%, according to PFAM) and 9,330 (69%, according to InterPro).

According to our results, the predicted secretomes of *C. cassiicola* CC_29 and *C. olivacea* CBS 114450 consisted of 1,327 (7.2%) and 920 (6.8%) proteins, respectively. The estimated fractions of repetitive DNA and transposable elements were slightly smaller in *C. cassiicola* CC_29 (2.04 Mbp, 4.6%) than in *C. olivacea* CBS 114450 (2.08 Mbp, 6.4%).

Our results suggested that the effector repertoires of both *C. cassiicola* CC_29 and *C. olivacea* CBS 114450 include proteins (CAZymes, proteases, lipases, and effectors) with potential to interact with plant substrates (Figure 2a) and enzymes associated with secondary metabolites (Figure 2b). Compared to the genome of *C. olivacea* CBS 114450, the genome of *C. cassiicola* CC_29 was richer in proteins that shared homology to CAZymes, proteases, lipases, and effectors. For example, there were 935 putative CAZymes for *C. cassiicola* CC_29 and 666 for *C. olivacea* CBS 114450. Out of those, 379 (40% in *C. cassiicola* CC_29) and 263 (39% in *C. olivacea*) were predicted to have a signal peptide. CAZymes are crucial during carbon acquisition and metabolism (Lo Presti et al., 2015). They are usually divided into six classes: glycoside hydrolases (GHs), carbohydrate esterases (CEs), glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate-binding modules (CBM), and auxiliary activities (AAs) (Lopez et al., 2018). In our genomes, the most represented class of CAZymes was GH, with 371 (40%) proteins in the *C. cassiicola* CC_29 genome and 290 (44%) in the *C. olivacea* CBS 114450 genome. The abundance of cell wall degrading enzymes (such as GHs) in the two genomes we described herein indicate that CAZymes may play a significant role during the interaction of those two fungi with their hosts during infection.

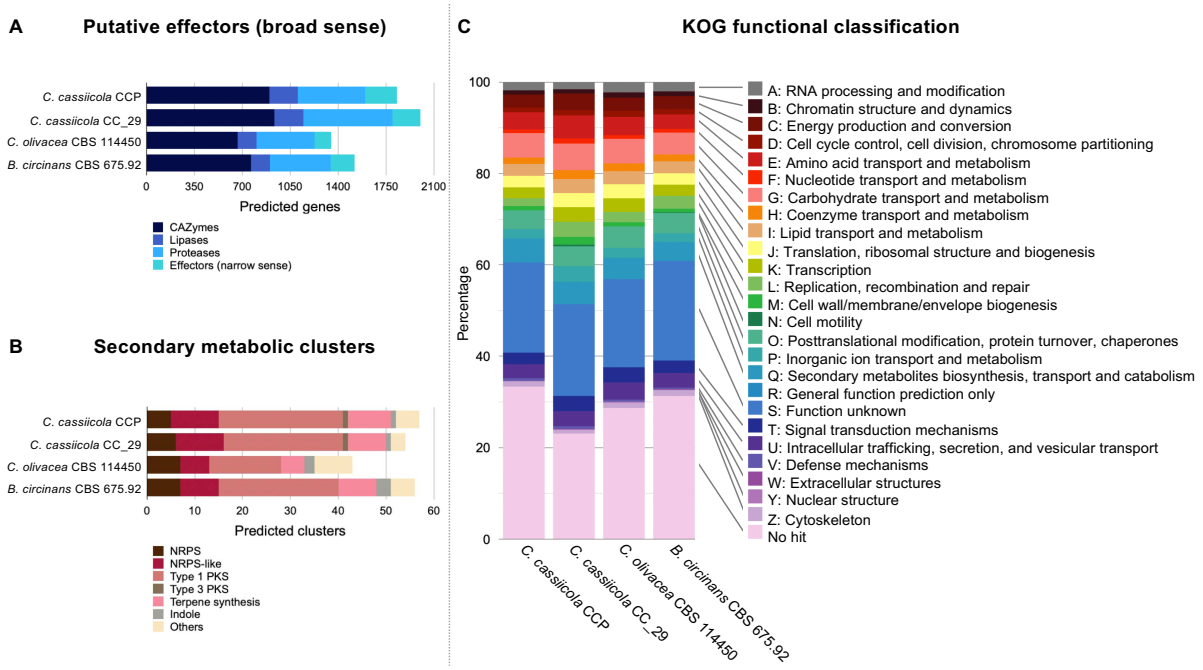


Figure 2. **Genome annotation of *Corynespora cassiicola* CC_29 and *C. olivacea* CBS 114450.** Annotations of *C. cassiicola* CCP and *Byssothecium circinans* CBS 675.92 are shown for comparative purpose. **A.** Number of predicted genes encoding groups of putative effectors in a broad sense. **B.** Number of gene clusters involved in the secondary metabolism. **C.** Distribution of KOG functional categories among protein-encoding genes. Names of each KOG category definition (A to Z) are provided on the right, numbers of genes in each category are displayed as percentage. Color codes as indicated.

The plant cuticle consists of cutin and wax and acts as a barrier against biotic and abiotic stresses (Ziv et al., 2018). To overcome this obstacle, plant pathogens secrete lipases and cutinases in order to promote the hydrolysis of fatty acids and enable the penetration into the host cell (Ohm et al., 2012). We predicted six cutinases each in the genomes of *C. cassiicola* CC_29 and *C. olivacea* CBS 114450. We also annotated putative lipases in both genomes. In the genome of *C. cassiicola* CC_29, there were 56 annotated lipases (27% out of a total of 211) that were predicted to have a signal peptide. Meanwhile, in the genome of *C. olivacea* CBS 114450, there were a total of 139 putative lipases (35% of which were predicted to have a signal peptide).

During infections of plant pathogens, proteases play roles in regulatory processes of several virulence factors; additionally, secreted proteases take part in preventing recognition by plant immune system and deactivating host defense responses (Figaj et al., 2019). We annotated 652 proteases in the *C. cassiicola* CC_29 genome and 421 in the *C. olivacea* CBS 114450 genome. Out of those, 128 (20%; *C. cassiicola* CC_29 genome) and 81 (19%; *C. olivacea* CBS 114450 genome) were predicted to have a signal peptide. Using a machine learning approach, we identified the effectors in a narrower sense from the predicted secretomes. There were 202 predicted narrow sense effectors in the genome of *C. cassiicola* CC_29 and 125 in the *C. olivacea* CBS 144450 genome.

Genes involved in the secondary metabolism also play roles during fungal development and establishment of interactions with other organisms (Keller, 2019). Our predictions revealed 54 genes associated with secondary metabolite pathways in the genome of *C. cassiicola* CC_29 and 43 in the genome of *C. olivacea* CBS 114450 (Figure 2b). In the genome of *Corynespora cassiicola* CC_29, there were genes that encoded for non-ribosomal peptide synthases (NRPS) (six genes), NRPS-like (ten), type 1 polyketide synthases (PKS) (25), type 3 PKS (one), terpene synthases (eight), and indoles (one). Similarly, the genome of *Corynespora olivacea* CBS 114450 contained genes that encoded for NRPS (seven genes), NRPS-like (six), type 1 PKS (15), terpene synthases (five), and indoles (two). When compared to the other orders within Dothideomycetes, Pleosporales species are recognized for their richness of gene clusters that are associated with secondary metabolites, especially PKS (Ohm et al., 2012).

The genomes of *Corynespora cassiicola* CC_29 and *C. olivacea* CBS 114450 shared a similar content of KOG functional annotation (Figure 2c). The main difference was in category P (protein encoding-genes in inorganic ion transport and metabolism), followed by category L (replication, recombination and repair). In category P, we annotated 639 (3.5%) genes in the genome of *C. cassiicola* CC_29, which corresponds to more than twice the number of genes (289; 2.1%) annotated in the genome of *C. olivacea* CBS 114450. In category L, we annotated 607 (3.3%) genes in the genome *C. cassiicola* CC_29 against 308 (2.3%) genes in the genome of *C. olivacea* CBS 114450.

For comparative purposes, we applied our full effector annotation pipeline and KOG functional classification to two previously characterized genomes: *C. cassiicola*

CCP (Lopez et al., 2018) and *B. circinans* CBS 675.92 (Haridas et al., 2020). *Corynespora cassiicola* CC_29 had the largest number of annotated effectors (in a broad sense), followed by *C. cassiicola* CCP (Figure 2a). *Corynespora olivacea* CBS 114450 had the lowest number of effectors among the four genomes. Isolates of *C. cassiicola* displayed the highest number of secondary metabolism gene clusters and *C. olivacea* CBS 114450, the lowest (Figure 2b). The richness in the effector repertoire of *C. cassiicola* compared to other species of Pleosporales has been attributed to a combination of factors, such as expansion in effector encoding-genes, wider host range, and multiple lifestyles (Lopez et al., 2018). The difference of putative effector repertoires between the two isolates of *C. cassiicola* (CC_29, this work; CCP, Lopez et al., 2018) likely resulted from a combination of acquisition/loss of accessory effectors during the species evolution (Gao et al., 2020; Lopez et al., 2018), geographic origin, and host species (CC_29 was isolated from soybean leaves collected in Brazil and CCP from rubber tree in the Philippines).

In summary, we presented the genome sequences of *C. cassiicola* CC_29, isolated from soybean leaves from Brazil, and *C. olivacea* CBS 114450. We performed a complete functional annotation and *in silico* mining of the effector repertoire for both genomes. The genome of *C. cassiicola* CC_29 contributes to our knowledge about virulence-associated genes of this emerging plant pathogen and also lays a foundation for the understanding of the plant-host interactions in a soybean-*Corynespora cassiicola* pathosystem. Using whole-genome data, we showed that *C. olivacea* is phylogenetically distant from its congener *C. cassiicola*. The genome of *C. olivacea* CBS 114450 provides useful information for future studies aimed at resolving the correct classification.

DATA ACCESSIBILITY

The whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession numbers JAGQEL000000000 for *Corynespora cassiicola* CC_29 and JAGQEM000000000 for *Corynespora olivacea* CBS 114450. The versions described in this paper are JAGQEL010000000 and JAGQEM010000000, respectively.

Acknowledgement

This work was supported by The Minas Gerais State Foundation of Research Aid – FAPEMIG (grant number APQ-00150-17) and by The National Council of Scientific and Technological Development – CNPq (fellowship number PQ 302336/2019-2) to LOO. TCSD received student fellowships from the CAPES Foundation (PROEX – 0487 No. 1684083) from Mar/2017 to Jul/2018 and CNPq (GM/GD 142400/2018-1) from Aug/2018 to Apr/2021. HVSR was supported by a PD fellowship from the São Paulo Research Foundation – FAPESP (2018/04555-0).

Conflict of interest

The authors have declared that no competing interests exist.

REFERENCES

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., Weber, T., 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 47, W81–W87.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Buchfink, B., Xie, C., Huson, D.H., 2014. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.

Crous, P.W., Summerell, B.A., Shivas, R.G., Romberg, M., Mel'nik, V.A., Verkley, G.J.M., Groenewald, J.Z., 2011. Fungal planet description sheets: 92-106. Persoonia Mol. Phylogeny Evol. Fungi 27, 130–162.

Dixon, L.J., Schlub, R.L., Pernezny, K., Datnoff, L.E., 2009. Host specialization and phylogenetic diversity of *Corynespora cassiicola*. Phytopathology 99, 1015–1027.

Doyle, J.J., Doyle, J.L., 1990. Isolation of plant DNA from fresh tissue. Focus (Madison). 12, 13–15.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., Finn, R.D., 2019. The Pfam protein families database in 2019. Nucleic Acids Res. 47, D427–D432.

Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 1–14.

Figaj, D., Ambroziak, P., Przepiora, T., Skorko-Glonek, J., 2019. The role of proteases in the virulence of plant pathogenic bacteria. Int. J. Mol. Sci. 20.

Fischer, M., Pleiss, J., 2003. The Lipase Engineering Database: A navigation and analysis tool for protein families. Nucleic Acids Res. 31, 319–321.

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., Smit, A.F., 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. U. S. A. 117, 9451–9457.

Gao, S., Zeng, R., Xu, L., Song, Z., Gao, P., Dai, F., 2020. Genome sequence and spore germination-associated transcriptome analysis of *Corynespora cassiicola* from cucumber. BMC Microbiol. 20, 1–20.

Grigoriev, I. V, Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., Shabalov, I., 2013. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. 42, D699.

Hane, J.K., Lowe, R.G.T., Solomon, P.S., Tan, K.C., Schoch, C.L., Spatafora, J.W., Crous, P.W., Kodira, C., Birren, B.W., Galagan, J.E., Torriani, S.F.F., McDonald, B.A., Oliver, R.P., 2007. Dothideomycete-plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. Plant Cell 19, 3347–3368.

Haridas, S., Albert, R., Binder, M., Bloem, J., LaButti, K., Salamov, A., Andreopoulos, B., Baker, S.E., Barry, K., Bills, G., Bluhm, B.H., Cannon, C., Castanera, R., Culley, D.E., Daum, C., Ezra, D., González, J.B., Henrissat, B., Kuo, A., Liang, C., Lipzen, A., Lutzoni, F., Magnuson, J., Mondo, S.J., Nolan, M., Ohm, R.A., Pangilinan, J., Park, H.J., Ramírez, L., Alfaro, M., Sun, H., Tritt, A., Yoshinaga, Y., Zwiers, L.H., Turgeon, B.G., Goodwin, S.B., Spatafora, J.W., Crous, P.W., Grigoriev, I. V., 2020. 101 Dothideomycetes genomes: A test case for predicting lifestyles and emergence of pathogens. Stud. Mycol. 96, 141–153.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., Von Mering, C., Bork, P., 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol. Biol. Evol. 34, 2115–2122.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., Von Mering, C., Bork, P., 2019. EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47, D309–D314.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: Genome-scale protein function classification. Bioinformatics 30, 1236–1240.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Keller, N.P., 2019. Fungal secondary metabolism: regulation, function and drug discovery. Nat. Rev. Microbiol. 17, 167–180.

Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E.L.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. 305, 567–580.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., Kahmann, R., 2015. Fungal effectors and plant susceptibility. Annu. Rev. Plant Biol. 66, 513–545.

Looi, H.K., Toh, Y.F., Yew, S.M., Na, S.L., Tan, Y.C., Chong, P.S., Khoo, J.S., Yee, W.Y., Ng, K.P., Kuan, C.S., 2017. Genomic insight into pathogenicity of dematiaceous fungus *Corynespora cassiicola*. PeerJ 5, 1–28.

Lopez, D., Ribeiro, S., Label, P., Fumanal, B., Venisse, J.S., Kohler, A., Oliveira, R.R., Labutti, K., Lipzen, A., Lail, K., Bauer, D., Ohm, R.A., Barry, K.W., Spatafora, J., Grigoriev, I. V., Martin, F.M., Pujade-Renaud, V., 2018. Genome-wide analysis of *Corynespora cassiicola* leaf fall disease putative effectors. Front. Microbiol. 9, 1–21.

Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., Gurevich, A., 2018. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics 34, i142–i150.

Molina, J.P.E., Paul, P.A., Amorim, L., da Silva, L.H.C.P., Siqueri, F. V., Borges, E.P., Campos, H.D., Venancio, W.S., Meyer, M.C., Martins, M.C., Balardin, R.S., Carlin, V.J., Grigolli, J.F.J., Belufi, L.M. d. R., Nunes Junior, J., Godoy, C. V., 2019. Effect of target spot on soybean yield and factors affecting this relationship. Plant Pathol. 68, 107–115.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.

Ohm, R.A., Feau, N., Henrissat, B., Schoch, C.L., Horwitz, B.A., Barry, K.W., Condon, B.J., Copeland, A.C., Dhillon, B., Glaser, F., Hesse, C.N., Kosti, I., LaButti, K., Lindquist, E.A., Lucas, S., Salamov, A.A., Bradshaw, R.E., Ciuffetti, L., Hamelin, R.C., Kema, G.H.J., Lawrence, C., Scott, J.A., Spatafora, J.W., Turgeon, B.G., de Wit, P.J.G.M., Zhong, S., Goodwin, S.B., Grigoriev, I. V., 2012. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. PLoS Pathog. 8.

Petersen, T.N., Brunak, S., Von Heijne, G., Nielsen, H., 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.

Rawlings, N.D., Barrett, A.J., Bateman, A., 2012. MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 40, 343–350.

Rouxel, T., Grandalbert, J., Hane, J.K., Hoede, C., Wouw, A.P. Van de, 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by repeat-induced point mutations. Nat. Commun. 2, 202.

Schoch, C.L., Crous, P.W., Groenewald, J.Z., Boehm, E.W.A., Burgess, T.I., de Gruyter, J., de Hoog, G.S., Dixon, L.J., Grube, M., Gueidan, C., Harada, Y., Hatakeyama, S., Hirayama, K., Hosoya, T., Huhndorf, S.M., Hyde, K.D., Jones, E.B.G., Kohlmeyer, J., Kruys, Å., Li, Y.M., Lücking, R., Lumbsch, H.T., Marvanová, L., Mbatchou, J.S., McVay, A.H., Miller, A.N., Mugambi, G.K., Muggia, L., Nelsen, M.P., Nelson, P., Owensby, C.A., Phillips, A.J.L., Phongpaichit, S., Pointing, S.B., Pujade-Renaud, V., Raja, H.A., Plata, E.R., Robbertse, B., Ruibal, C., Sakayaroj, J., Sano, T., Selbmann, L., Shearer, C.A., Shirouzu, T., Slippers, B., Suetrong, S., Tanaka, K., Volkmann-Kohlmeyer, B., Wingfield, M.J., Wood, A.R., Woudenberg, J.H.C., Yonezawa, H., Zhang, Y., Spatafora, J.W., 2009. A class-wide phylogenetic assessment of Dothideomycetes. Stud. Mycol. 64, 1–15.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E.M., 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.

Snelders, N.C., Kettles, G.J., Rudd, J.J., Thomma, B.P.H.J., 2018. Plant pathogen effector proteins as manipulators of host microbiomes? Mol. Plant Pathol. 19, 257–259.

Sonah, H., Deshmukh, R.K., Bélanger, R.R., 2016. Computational prediction of effector proteins in fungi: Opportunities and challenges. Front. Plant Sci. 7, 1–14.

Sperschneider, J., Dodds, P.N., Gardiner, D.M., Singh, K.B., Taylor, J.M., 2018. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Mol. Plant Pathol. 19, 2094–2110.

Stanke, M., Morgenstern, B., 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33, 465–467.

Sumabat, L.G., Kemerait, R.C., Brewer, M.T., 2018. Phylogenetic diversity and host specialization of *Corynespora cassiicola* responsible for emerging target spot disease of cotton and other crops in the southeastern United States. Phytopathology 108, 892–901.

Tanaka, K., Hirayama, K., Yonezawa, H., Sato, G., Toriyabe, A., Kudo, H., Hashimoto, A., Matsumura, M., Harada, Y., Kurihara, Y., Shirouzu, T., Hosoya, T., 2015. Revision of the Massarineae (Pleosporales, Dothideomycetes). Stud. Mycol. 82, 75–136.

Verma, S., Gazara, R.K., Nizam, S., Parween, S., Chattopadhyay, D., Verma, P.K., 2016. Draft genome sequencing and secretome analysis of fungal phytopathogen

Ascochyta rabiei provides insight into the necrotrophic effector repertoire. Sci. Rep. 6, 1–14.

Voglmayr, H., Jaklitsch, W.M., 2017. *Corynespora, Exosporium* and *Helminthosporium* revisited – New species and generic reclassification. Stud. Mycol. 87, 43–76.

Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. Genome Biol. 20, 1–13.

Zeiner, C.A., Purvine, S.O., Zink, E.M., Paša-Tolić, L., Chaput, D.L., Haridas, S., Wu, S., LaButti, K., Grigoriev, I. V., Henrissat, B., Santelli, C.M., Hansel, C.M., 2016. Comparative analysis of secretome profiles of manganese(II)-oxidizing *Ascomycete* fungi. PLoS One 11.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P.K., Xu, Y., Yin, Y., 2018. DbCAN2: A meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 46, W95–W101.

Ziv, C., Zhao, Z., Gao, Y.G., Xia, Y., 2018. Multifunctional roles of plant cuticle during plant-pathogen interactions. Front. Plant Sci. 9, 1–8.

Supplementary material

Genome sequences and in silico effector mining of *Corynespora cassiicola* CC_29 and *Corynespora olivacea* CBS 114450

Table S1. Detailed information on all fungal species used in this study.

Species / Isolate	Family	Order	Reference
Ascochyta rabiei ArDII	*Didymellaceae*	Pleosporales	Verma et al., 2016
Bimuria novae-zelandiae CBS 107.79	*Montagnulaceae*	Pleosporales	Haridas et al., 2020
Byssothecium circinans CBS 675.92	*Massarinaceae*	Pleosporales	Haridas et al., 2020
Clathrospora elynae CBS 161.51	*Diademaceae*	Pleosporales	Haridas et al., 2020
Cochliobolus heterostrophus C5	*Pleosporaceae*	Pleosporales	Ohm et al., 2012
Cochliobolus sativus ND90Pr	*Pleosporaceae*	Pleosporales	Ohm et al., 2012
Corynespora cassiicola CCP	*Corynesporascaceae*	Pleosporales	Lopez et al., 2018
Corynespora cassiicola CC_29	*Corynesporascaceae*	Pleosporales	This work
Corynespora olivacea CBS 114450	*Massarinaceae*	Pleosporales	This work
Didymella exigua CBS 183.55	*Didymellaceae*	Pleosporales	Haridas et al., 2020
Hysterium pulicare	*Hysteriaceae*	Hysteriales	Ohm et al., 2012
Karstenula rhodostoma CBS 690.94	*Montagnulaceae*	Pleosporales	Haridas et al., 2020
Lentithecium fluviatile	*Lentitheciaceae*	Pleosporales	Haridas et al., 2020
Leptosphaeria maculans	*Leptosphaeriaceae*	Pleosporales	Rouxel et al., 2011
Massarina eburnea CBS 473.64	*Massarinaceae*	Pleosporales	Haridas et al., 2020
Paraconiothyrium sporulosum AP3s5-JAC2a	*Montagnulaceae*	Pleosporales	Zeiner et al., 2016
Phoma tracheiphila IPT5	Incertae sedis	Pleosporales	Haridas et al., 2020
Setosphaeria turcica Et28A	*Pleosporaceae*	Pleosporales	Ohm et al., 2012
Stagonospora nodorum SN15	*Phaeosphaeriaceae*	Pleosporales	Hane et al., 2007
Trematosphaeria pertusa CBS 122368	*Trematosphaeriaceae*	Pleosporales	Haridas et al., 2020

Chapter 2

Evolutionary history of the Necrosis and Ethylene-inducing peptide 1-like protein (NLP) superfamily across the Dothideomycetes class of fungi

Evolutionary history of the Necrosis and Ethylene-inducing peptide 1-like protein (NLP) superfamily across the Dothideomycetes class of fungi

Thaís Carolina da Silva Dal'Sasso[1], Hugo Vianna Silva Rody[2], Luiz Orlando de Oliveira[1]

[1]Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Viçosa, Viçosa, Brazil
[2]Departamento de Genética, Universidade de São Paulo/Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, Brazil

Corresponding author: Luiz Orlando de Oliveira
Email: lorlando@ufv.br

ABSTRACT

Necrosis and Ethylene-inducing peptide 1-like proteins (NLPs) are broadly distributed across bacteria, fungi and oomycetes. Cytotoxic NLPs are usually secreted into the host apoplast where they can induce cell death and trigger plant immune responses in eudicots. NLP superfamily size is highly variable, but usually small among Dothideomycetes. We accessed the genomic resources of 33 species from five orders of Dothideomycetes to investigate the evolutionary history of the NLPs in the class. Phylogenetic approaches searched for biased patterns of NLP gene evolution and aimed to provide a phylogenetic framework for the cytotoxic activities of NLPs. NLP superfamily sizes were higher among necrotrophic fungi and composed mostly by NLP effectors. Members of the NLP1 family (type I NLPs) were predominant (89%) over members of the NLP2 family (type II NLPs). The NLP1 family split into three subfamilies (NLP1.1, NLP1.2, NLP1.3). The NLP1.1 subfamily was broadly distributed across those Dothideomycetes species that had at least one NLP member. There were strong concordances between the phylogenomics of Dothideomycetes and phylogenies of individual NLP1 subfamilies and NLP2 family. While gene losses seen to have shaped the evolutionary history of NLP2 family, ancient gene duplications followed by descent with modification shaped the NLP1 family. The strongest cytotoxic activities were recorded on NLPs of the NLP1.1 subfamily, suggesting that cytotoxicity was associated with preferential NLP gene retention of members of that subfamily.

Keywords: Pleosporales, cytotoxicity, effectors, biased gene evolution

INTRODUCTION

Necrosis and Ethylene-inducing peptide 1-like proteins (NLPs) are a superclass of proteins present in bacteria, fungi and oomycetes (Gijzen and Nürnberger, 2006). In general, NLPs are small proteins of about 24 kDa (Pemberton and Salmond, 2004) that exhibit cytotoxic activity with cell death-inducing properties and trigger plant immune response in eudicot plants, but not in monocots (Gijzen and Nürnberger, 2006; Pemberton and Salmond, 2004; Seidl and Van den Ackerveken, 2019). The first purified NLP was the Necrosis- and Ethylene-inducing peptide 1 (Nep1) from *Fusarium oxysporum* f. sp. *erythroxyli,* a protein capable of inducing necrosis and the production of ethylene when applied to *Erythroxylum coca* (coca) (Bailey, 1995).

A number of evidence-based studies indicated that NLPs of plant-associated fungi may play a role as virulence factors during the early stages of infection and disease development. The removal of cytotoxic NLP-enconding gene resulted in fungal mutants with decreased virulence (Levin et al., 2019; Santhanam et al., 2013). On other instances, the overexpression of a cytotoxic NLP-enconding gene increased the virulence of the mutant (Amsellem et al., 2002). The role of NLPs in the infection process may not be iqually important for all phytopathogens. For example, the virulence of *Magnaporthe oryzae* (Fang et al., 2017) and *Botrytis elliptica* (Staats et al., 2007) mutants did not change when their NLP-enconding genes were removed. Through functional analyses, some members of the prolific NLP superfamilies of *Diplodia seriata* (Cobos et al., 2019) and *Neofusicoccum parvum* (Pour et al., 2020) were shown to exhibit varying levels of citotoxicity. Contrary to cytolytic NLPs, the noncytolytic NLPs cannot permeabilize the plant membrane but retain the capability of triggering plant-immune responses; their function are yet to be characterized (Cabral et al., 2012; Dong et al., 2012; Ottmann et al., 2009; Qutob et al., 2006).

The defining characteristic of an NLP is its NPP1 domain (Pfam PF05630) (Fellbrich et al., 2002). This domain contains a highly conserved, seven amino acid long motif (GHRHDWE) that are involved in cation binding (Böhm et al., 2014; Oome et al., 2014; Ottmann et al., 2009). Consistent with their role as proteins of the secretory pathway, the vast majority of NLPs have an N-terminal signal peptide (Oome and Van Den Ackerveken, 2014; Seidl and Van den Ackerveken, 2019). In the N-terminal half of NLPs, there are conserved cysteine residues that are able to form disulfide bridges, which seem to be essential for protein stabilization and necrosis induction (Fellbrich et

al., 2002; Ottmann et al., 2009).

Based on the number of the conserved cysteines, NLPs have been classified into four main types. Type I NLPs have two conserved cysteine residues and are the most widespread type, occurring in fungi, bacteria, and oomycetes (Gijzen and Nürnberger, 2006; Oome and Van Den Ackerveken, 2014; Seidl and Van den Ackerveken, 2019). Type Ia also have two cysteine but differs from type I by the occurrence of distinct amino acid substitutions; they are found among oomycetes (Oome and Van Den Ackerveken, 2014). Type II NLPs have four conserved cysteines, which are responsible for the formation of two disulfide bridges; type II NLPs occur less frequently, mostly in bacteria and fungi (Gijzen and Nürnberger, 2006). Type III NLPs exhibit the least conserved amino acid sequence, most of them carry six cysteine residues that are able to form three disulfde bridges; they occur in ascomycetes and some bacteria (Seidl and Van den Ackerveken, 2019). The occurrence of types I, II, and III NLPs in Ascomycota suggests that NLPs originated within this phylum; afterwards, horizontal gene transfer (HGT) allowed NLPs to reach distantly-related taxa (Oome and Van Den Ackerveken, 2014).

A recent study surveyed the taxonomic distribution of NLPs in reference proteomes of about 10 thousand species of bacteria, fungi, and oomycetes; about 500 species contained each at least one NLP (Seidl and Van den Ackerveken, 2019). The most stricking features that bring together most of the species that do contain NLPs are their taxonomic placement within Ascomycota (60%, 211 of 360 proteomes), plant pathogenicity (some of the most obnoxious phytopathogens do contain NLPs), and lifestyles (specially biotrophic, hemibiotrophic and necrotrophic microbes).

The Dothideomycetes comprises the largest and phylogenetically most-diverse class within Ascomycota, with an estimated number of members that may reach up to 19 thousand species (Hyde et al., 2013; Kirk et al., 2008). They occur across diverse habitats, including extreme environmental conditions and their lifestyles are very diverse (Haridas et al., 2020; Hyde et al., 2013; Schoch et al., 2009). Dothideomycetes diverged from other Ascomycota around 366 million years ago (mya) (Liu et al., 2017). Currently, the class comprises 23 orders (Haridas et al., 2020) that are believed to have evolved in the range between 100 and 220 mya (Liu et al., 2017). The Pleosporales, the Capnodiales, and the Botryosphaeriales are three of the largest orders within the Dothideomycetes; each of those orders contains a large number of families, which in turn hold some of the most destructive genera of plant pathogens to

cereal crops, trees, and dicots (Haridas et al., 2020; Ohm et al., 2012; Schoch et al., 2009).

The diversification of the NLP superfamily across Dothideomycetes is complex and far from being understood. The superfamily may not have reached the extreme large size found in Oomycetes, in which up to 100 members have been documented (Seidl and Van den Ackerveken, 2019). Botryosphaeriales, however, is among the most NLP-rich orders among Dothideomycetes. *Neofusicoccum parvum* with six NLPs (Pour et al., 2020) and *Diplodia seriata* with four NLPs (Cobos et al., 2019) typify the NLP superfamily size in this order. Pleosporales is an order in which the NLP superfamily size is consistently small, with most species analyzed to date having one or two members (Oome and Van Den Ackerveken, 2014; Seidl and Van den Ackerveken, 2019). There are rare cases in which three NLPs have been documented in Pleosporales, such as in the plant pathogens *Didymela exigua* (Oome and Van Den Ackerveken, 2014) and *Cochliobolus sativus* (Seidl and Van den Ackerveken, 2019).

With incredible ecological and morphological diversities and economical importance as plant pathogens, the Dothideomycetes have raised the interest of genomic research in the recent decade. Currently, genomes of about 90 genera of Dothideomycetes are available in public repositories, most of which are from plant pathogens and plant-associated species (Haridas et al., 2020; Ohm et al., 2012). Those genomic resources allow for the investigation of the evolutionary history of NLPs across that class.

We began our study by building up a robust molecular phylogenetic framework based on a set of 1,858 single copy orthologs proteins of 33 species of Dothideomycetes. To account for ecological diversity within the class, we included species of five lifestyles (biotrophic, hemibiotrophic, necrotrophic, saprotrophic, and ectomycorrhizal) sorted out among plant pathogens, plant-associated species, and non-plant associated species. Built independently from the NLP evolutionary history, our framework revealed the phylogenomic relationships among the study species and should shed light into the evolutionary history of NLPs across Dothideomycetes. Next, we reconstructed the phylogenetic relationships among the predicted type I and type II NLPs that we had retrieved from those genomic resources and questioned how ubiquitous and diverse the NLP superfamily became across that class of fungi. If NLP gene evolution took place in an NLP type-independent manner, both type I NLPs and type II NLPs should show similar distribution across the NLPs and species

phylogenies. Otherwise, if biased patterns drove NLP gene evolution, the NLP phylogeny will be imbalanced, as the type that has been favored throughout evolution should have become more widespread. The negatively selected type will be rare; likely vanishing from certain species subclades. Subsequently, we explored how the processes that drove NLP gene evolution took place at the lower ranks of the NLP phylogenetic hierarchy. To answer our questions, we build independent phylogenies, one for each of the major NLP subclades we had recovered previously. Then, we contrasted the topology of those trees with the phylogenetic framework of Dothideomycetes. Finally, our study provided a context to explore whether the cytotoxic activities reported through functional analyses are associated with NLP gene evolution and phylogenetic relationships among species. If preferential gene retention patterns maintained functional paralogs over time, cytotoxic activities will tend to exhibit some phylogenetic signal. This investigation shed light on the likely mechanisms that contributed to the evolution of NLPs in Dothideomycetes.

MATERIAL AND METHODS

Data assembly

We assembled protein sequence data from 33 species of Dothideomycetes. From MycoCosm (Grigoriev et al., 2013), we retrived protein sequences of 32 species of Dothideomycetes and one additional species of Eurotiomycetes (*Aspergillus nidulans*). Our analyses also included genomic data from *Corynespora olivacea* isolate CBS 114450 (Chapter 1).

Orthogroups and phylogenomic analysis

OrthoFinder v2.3.3 (Emms and Kelly, 2015) established orthologous relationships among members of the Dothideomycetes (Table 1). OrthoFinder calculated length and phylogenetic distance-normalized scores from an all-versus-all alignment using DIAMOND v0.9.24 (Buchfink et al., 2014) and identified reciprocal best normalized hits. Normalized scores above default thresholds were assigned to the orthogroup graph and subjected to the Markov clustering analysis in order to assume ortholog groups (orthogroups).

We used a Maximum-likelihood phylogeny approach to uncovered the phylogenetic relationships among species of the Dothideomycetes. To reconstruct the phylogeny, we prepared a dataset with the single copy ortholog (SCO) proteins identified by OrthoFinder. Proteins were aligned using the L-INS-I method as implemented in MAFFT v7.453 (Katoh and Standley, 2013). trimAL v1.4.rev22 (Capella-Gutiérrez et al., 2009) trimmed the alignments, with the parameters “-gt 0.7”, “-cons 80”, and “-w 5”. After trimming and concatenation, we obtained dataset 1.

We subjected dataset 1 as input to IQ-TREE v1.6.11 (Nguyen et al., 2015) to estimate both the best evolutionary model and the phylogenomic tree. According to the Bayesian information criteria (BIC), IQ-TREE standard model selection indicated LG+F+I+G4 as the best fit model. Subsequently, we carried out a Maximum-Likelihood phylogenomic tree using 1,000 ultrafast bootstrap replicates and ten independent runs. *Aspergillus nidulans* was set as outgroup. The tree was visualized in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Table 1. Detailed information on all fungal species used in this study.

Species/Isolate	Alias	Order/Class	Lifestyle	Genome Size (Mb)	# Genes	Reference	# Predicted NLP Genes	# Predicted NLP Effectors
Ascochyta rabiei ArDII	Ascra	Pleosporales/Dothideomycetes	Necrotrophic	34.65	10,596	Verma et al., 2016	2	2
Aspergillus nidulans	Aspni	Eurotiales/Eurotiomycetes	Saprotrophic	30.48	10,608	Arnaud et al., 2012	2	2
Baudoinia compniacensis UAMH 10762 (4089826)	Bauco	Capnodiales/Dothideomycetes	Saprotrophic	21.88	10,513	Ohm et al., 2012	0	0
Bimuria novae-zelandiae CBS 107.79	Bimnz	Pleosporales/Dothideomycetes	Saprotrophic	78.18	16,681	Haridas et al., 2020	2	2
Botryosphaeria dothidea	Botdo	Botryosphaeriales/Dothideomycetes	Necrotrophic	43.50	14,998	Marsberg et al., 2017	6	0
Byssothecium circinans CBS 675.92	Bysci	Pleosporales/Dothideomycetes	Saprotrophic	49.29	15,785	Haridas et al., 2020	3	2
Cenococcum geophilum 1.58	Cenge	Mytilinidiales/Dothideomycetes	Ectomycorrhizal	17.75	14,748	Peter et al., 2016	0	0
Cercospora zeae-maydis	Cerzm	Capnodiales/Dothideomycetes	Necrotrophic	46.61	12,020	Haridas et al., 2020	1	1
Cladosporium fulvum	Clafu	Capnodiales/Dothideomycetes	Biotrophic	61.11	14,127	De Wit et al., 2012	1	0
Clathrospora elynae CBS 161.51	Clael	Pleosporales/Dothideomycetes	Saprotrophic	37.39	13,617	Haridas et al., 2020	1	1
Cochliobolus heterostrophus C5	Coche	Pleosporales/Dothideomycetes	Necrotrophic	36.46	13,336	Ohm et al., 2012	2	2
Cochliobolus sativus ND90Pr	Cocsa	Pleosporales/Dothideomycetes	Hemibiotrophic	34.42	12,250	Ohm et al., 2012	3	2
Corynespora cassiicola CCP	Corca	Pleosporales/Dothideomycetes	Necrotrophic	44.85	17,166	Lopez et al., 2018	3	3
Corynespora olivacea CBS 114450	Corol	Pleosporales/Dothideomycetes	Saprotrophic	32.32	13,501	Chapter 1	1	0
Didymella exigua CBS 183.55	Didex	Pleosporales/Dothideomycetes	Saprotrophic	34.39	12,394	Haridas et al. 2020	3	2
Diplodia seriata DS831	Dipse	Botryosphaeriales/Dothideomycetes	Necrotrophic	37.12	9,343	Morales-Cruz A et al., 2015	4	4
Dothistroma septosporum NZE10	Dotse	Capnodiales/Dothideomycetes	Hemibiotrophic	30.21	12,580	De Wit et al., 2012	0	0
Hysterium pulicare	Hyspu	Hysteriales/Dothideomycetes	Saprotrophic	38.43	12,352	Ohm et al., 2012	1	0
Karstenula rhodostoma CBS 690.94	Karrh	Pleosporales/Dothideomycetes	Saprotrophic	44.89	12,469	Haridas et al., 2020	1	1
Lentithecium fluviatile	Lenfl	Pleosporales/Dothideomycetes	Saprotrophic	54.69	16,742	Haridas et al., 2020	1	0
Leptosphaeria maculans	Lepma	Pleosporales/Dothideomycetes	Hemibiotrophic	44.89	12,469	Rouxel et al., 2011	2	1
Macrophomina phaseolina MS6	Macph	Botryosphaeriales/Dothideomycetes	Necrotrophic	48.88	13,806	Islam et al., 2012	4	3
Massarina eburnea CBS 473.64	Maseb	Pleosporales/Dothideomycetes	Saprotrophic	38.24	12,935	Haridas et al., 2020	1	1
Neofusicoccum parvum UCRNP2	Neopa	Botryosphaeriales/Dothideomycetes	Necrotrophic	42.59	10,366	Blanco-Ulate et al., 2013	6	5
Paraconiothyrium sporulosum AP3s5-JAC2a	Parsp	Pleosporales/Dothideomycetes	Saprotrophic	38.46	14,745	Zeiner et al., 2016	2	2
Phoma tracheiphila IPT5	Photr	Pleosporales/Dothideomycetes	Necrotrophic	34.24	13,209	Haridas et al., 2020	1	1
Pseudocercospora (Mycosphaerella) fijiensis	Mycfi	Capnodiales/Dothideomycetes	Hemibiotrophic	74.14	13,107	Isaza et al, 2016	2	1
Rhytidhysteron rufulum	Rhyru	Hysteriales/Dothideomycetes	Saprotrophic	40.18	12,117	Ohm et al., 2012	1	0
Saccharata proteae CBS 121410	Sacpr	Botryosphaeriales/Dothideomycetes	Saprotrophic	31.43	9,234	Haridas et al., 2020	0	0
Septoria musiva SO2202	Sepmu	Capnodiales/Dothideomycetes	Hemibiotrophic	29.35	10,233	Ohm et al., 2012	1	1
Septoria populicola	Seppo	Capnodiales/Dothideomycetes	Hemibiotrophic	33.19	9,739	Ohm et al., 2012	1	1
Setosphaeria turcica Et28A	Settu	Pleosporales/Dothideomycetes	Hemibiotrophic	43.01	12,028	Ohm et al., 2012	1	1
Stagonospora nodorum SN15	Stano	Pleosporales/Dothideomycetes	Necrotrophic	37.21	12,380	Hane et al., 2007	2	2
Trematosphaeria pertusa CBS 122368	Trepe	Pleosporales/Dothideomycetes	Saprotrophic	47.74	17,306	Haridas et al., 2020	1	0

Protein annotation and assemble of NLP homologues

The predicted proteomes of the set of 34 species were annotated using PfamScan with Pfam v32.0 (El-Gebali et al., 2019) and InterProScan v5.30.69 (Jones et al., 2014) with the following eight parameters: SMART-7.1, SUPERFAMILY-1.75, ProDom-2006.1, CDD-3.16, TIGRFAM-15.0, Pfam v31.0, Coils-2.2.1, and Gene3D-4.2.0.

The presence of a signal peptide (according to SignalP v4.1; Petersen et al., 2011) and the absence of transmembrane domains (according to TMHMM v2.0; (Krogh et al., 2001) defined the proteins that we predicted to be part of the secretome. Subsequently, TargetP (Emanuelsson et al., 2007) predicted the subcellular localization of the predicted secretome.

Finally, HMMER v3.2.1 (http://hmmer.org) predicted the NLP homologues (E-value < 0.001) using profile hidden markov models (HMMs) for NPP1 domain (PF05630) from the predicted proteomes. A protein was declared to be a NLP when the NPP1 domain was annotated by at least two out three softwares: PfamScan, InterproScan, and HMMER. Within the pool of NLPs, a given protein was considered to be a NLP effector when it passes the following three tests: (a) SignalP predicted it harbors a signal peptide, (b) TMHMM predicted it has no transmembrane domain, and (c) TargetP predicted it was part of a secretoy pathway. When a given NLP failed any of these three tests, it was regarded as a “non-effector NLP”. Finally, we assembled dataset 2, which contained the predicted NLPs (protein data) that were present in the set of 34 species.

NLP Bayesian phylogenies

Bayesian phylogenetic analysis based on dataset 2 allowed us to reconstruct the phylogenetic relationships of NLPs among the members of the Dothideomycetes.

Alingments were obtained using the L-INS-I method, as implemented in MAFFT v7.453. MAFFT L-INS-I allows for aligning a set of sequences containing sequences flanking around one alignable domain with high accuracy (Katoh and Standley, 2013). To find the best fit model, we used dataset 2 and its partitions as input to ProtTest v3.4.2 (Darriba et al., 2011). According to BIC, ProtTest selected WAG+I+G for dataset 2 and WAG+G for its partitions as the best evolutionary models. Subsequently, phylogenetic analyses were conducted on MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003), using two simultaneous runs, one cold chain and seven heated

chains in each run. During the runs, the number of generations was ten million and temperature parameter was 0.23. We also analyzed the phylogenetic relationships of NLPs among subclades we had obtained in the previous analysis. The number of generations was five million and temperature parameter was 0.25. Trees were sampled every 5000 generations and the first 250 trees were discarded as burn-in samples. Convergence was diagnosed with a) the standard deviation of split frequencies at the end of each run (below 1.5%; according to the output from MrBayes) and b) the Effective Sample Size (ESS) of each parameter (above 200 for all statistics) in Tracer v1.7.1 (Rambaut et al., 2018). For each analysis, a 50% majority-rule consensus tree of the two independent runs was obtained with posterior probabilities (PP) that were equal to bipartition frequencies. The phylogenetic trees were visualized in FigTree v1.4.4.

Survey on functional activities of NLPs

We surveyed the literature for functional analyses of NLPs across Ascomycota. We recorded the taxonomic placement of the study species, the NLP names attributed to the proteins, and the original outcome of the functional analyses (usually each study stated the activity as strong, weak, or absent). When the species was in Dothideomycetes, we included it in our phylogenetic approach. Species of other classes of fungi were not included in our phylogenies, but we carried out protein sequence alignments individually in order to assign the placement of each of their NLPs into our phylogenetic treatments.

RESULTS

Genome-wide phylogeny of Dothideomycetes

The set of 33 species (30 genera) of Dothideomycetes allowed OrthoFinder to uncover 1,858 SCO proteins (a concatenated set of 1,099,121 amino acids). The whole-genome data provided suport for the phylogenetic reconstruction within Dothideomycetes (Figure 1). The phylogenomic tree showed well-supported nodes (bootstrap values = 100, for all nodes). Overall, there were two major clades. The first major clade held all seven species of the order Capnodiales. The second major clade contained 26 species within the remaining four orders of Dothideomycetes

(Pleosporales, Hysteriales, Mytilinidiales, and Botryosphaeriales). Sister to Hysteriales, there was a clade that brought together 18 species of Pleosporales. The Pleosporales clade split further into three sub-clades. One of the sub-clades brought together nine species of distinct lifestyles (necrotrophs, hemibiotrophs, or saprotrophs). There was another sub-clade, in which all eight species were saprotrophs, including *Corynespora olivacea*. Finally, showing a sister relationship to the other sub-clades of Pleosporales, there were a sub-clade that contained *Corynespora cassiicola* as its only member.

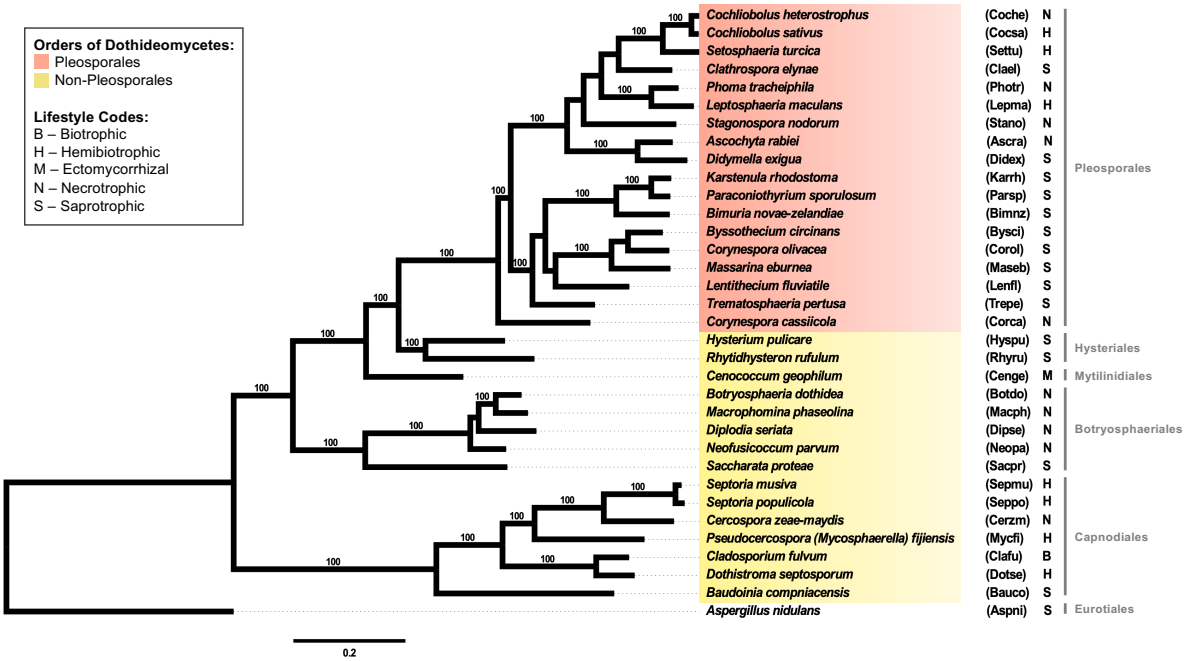


Figure 1. Phylogenomic relationships among 33 species of Dothideomycetes. The maximum likelihood consensus tree is based on a dataset of 1,858 single copy ortholog proteins, with *Aspergillus nidulans* (Eurotiomycetes) as outgroup. Species names are color-coded according to their placements within Dothideomycetes (Pleosporales or non-Pleosporales, as indicated). Five-letter codes indicate alias of each species name. Lifestyles are coded as indicated. Branch lengths are drawn to scale; nodal support values are given as local bootstraps above the branches. Scale bar corresponds to the expected number of substitutions per site.

Genome-wide identification of homologues of NLPs in Dothideomycetes

Apart from the NPP1 domain (PF05630), our pipeline predicted no other domain as part of the NLPs. Each NLP had a single copy of the NPP1 domain. Among the NLPs, our pipeline distinguished between two sets of proteins: (a) NLP effector and (b) non-effector NLPs. In the first set, the protein harbored a signal peptide, had no transmembrane domain, and was part of the secretory pathway. In the second set, despite the presence of the NPP1 domain, the protein failed to comply with any of the previous three requirements.

In general, NLP superfamily sizes among species of Dothideomycetes varied (Table 1). There were four non-Pleosporales species without any NLPs predicted in their genomes: two saprotrophs (*Baudoinia compniacensis* and *Saccharata proteae*), one hemibiotroph (*Dothistroma septosporum*), and one ectomycorrhizal (*Cenococcum geophilum*). Within the 29-remaining species of Dothideomycetes, NLP family sizes ranged from one (14 species) to six members (*Bothryosphaeria dothidea* and *Neofusicoccum parvum*, both of them are necrotrophic species within Botryosphaeriales). Fungal lifestyle seen to be associated with NLP superfamily size: The mean number of NLP family members reached the largest value among necrotrophic fungi (3.1 NLPs; 2.3 NLP effectors); then follows the hemibiotrophs (1.4 NLPs; 1.0 NLP effectors) and the saprotrophs (1.3 NLPs; 0.9 NLP effectors). The only biotrophic species in our analyses (*Cladosporium fulvum*) harbored a single non-effector NLP gene.

The general trend was that NLP effectors predominate over non-effector NLPs. The copy number of NLP effectors ranged from one (11 species) to five (*Neofusicoccum parvum*). However, the NLP families of seven species (*Bothyosphaeria dothidea, Cladosporium fulvum, Corynespora olivacea, Dothistroma septosporum, Lentithecium fluviatile, Rhytidhysteron rufulum* and *Trematosphaeria pertusa*) contained exclusively non-effector NLPs (from 1 to 6 members).

Broad sense phylogeny of NLPs

After carrying out a phylogenomic reconstruction within the set of 33 species of Dothideomycetes (Figure 1), we explored again the phylogenetic relationships among species considering protein data from the NLPs as the only source of information. Those four species that lacked NLPs did not take part in this reconstruction. The Bayesian phylogenetic tree contained a total of 62 NLPs from 29 species of Dothideomycetes, together with *A. nidulans* (Figure 2, Table 2). The phylogenomic tree showed well-supported nodes (PP > 95%, for most nodes). The number of conserved cysteine residues in the protein determined the placement of the NLP in the tree. There were two major clades. A major clade harbored 55 NLPs that contained two cysteine residues (Type I NLP); another major clade brought together seven NLPs that contained four cysteine residues (Type II NLP). According to OrthoFinder, each of the two major clades corresponded to an orthogroup. Additionally, the presence of both NLP effectors and non-effector NLPs are spread across the phylogeny.

Type I NLPs formed the most member-rich clade. It contained 54 NLPs from 29 species of Dothideomycetes and one NLP from *A. nidulans.* Among the type I NLPs, 38 (70%) were predicted to be effector NLPs. The type I NLP clade was split further into four sub-clades. One of the four sub-clades contained a highly differentiated NLP effector from *Paraconiothyrium sporulosum* (Parsp_876375) as the single member. Species composition among the remaining three sub-clades was very diverse. Among the 33 species, *Corynespora cassiicola* was the only one that was represented in all sub-clades by a single member only.

Meanwhile, Type II NLPs formed a very small clade. It contained seven NLPs, each of which from a species of Dothideomycetes and one NLP from *A. nidulans.* Among Type II NLPs, six (86%) were predicted to be effector NLPs.

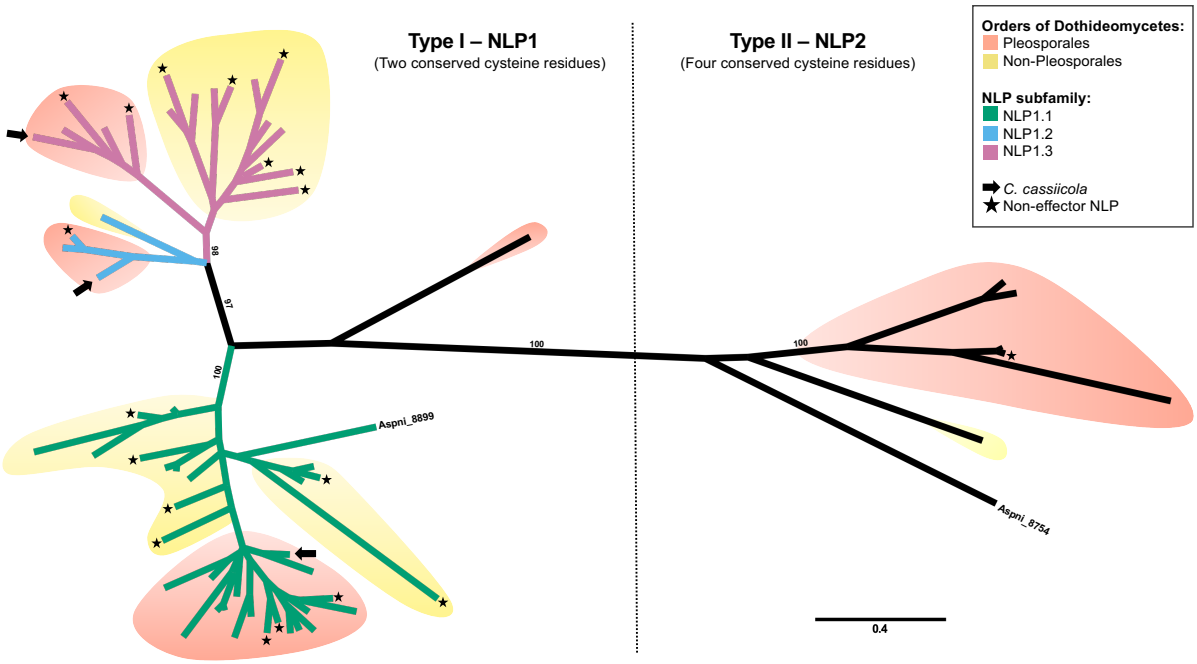


Figure 2. Phylogenetic relationships among 60 predicted Necrosis- and Ethylene-inducing Like proteins (NLPs) in Dothideomycetes. The dataset for the unrooted Bayesian phylogeny (consensus tree) comprised proteins that were 490 amino acids long. Two proteins from *Aspergillus nidulans* (Eurotiomycetes) are included for reference purpose. Clades are color-coded according their placements within Dothideomycetes (Pleosporales or non-Pleosporales, as indicated). Branches are color-coded according to the NLP1 subfamily, as indicated. Arrows indicate the phylogenetic placements of the three NLPs of *Corynespora cassiicola* isolate CCP. Along the phylogeny, stars indicate non-effector NLPs. A vertical bar splits the NLP tree into two sections, according to the number of conserved cysteine residues in the protein (Type I, two residues; Type II, four residues). Branch lengths are drawn to scale; nodal support values are given as posterior probabilities (when ≥ 95%) above the branches. Scale bar corresponds to the expected number of substitutions per site.

Table 2. Detailed information on NLPs from all species in Figure 2.

Alias	Protein ID according to MycoCosm	Order of Dothideomycetes	Type	Effector/ Non-effector	NLP (sub)family	Orthogroup ID according to OrthoFinder analysis
Ascra	6363	Pleosporales	I	Effector	NLP1.1	OG0000785
Ascra	9348	Pleosporales	I	Effector	NLP1.3	OG0000785
Aspni	8764	-	II	Effector	NLP2	OG0010037
Aspni	8899	-	I	Effector	NLP1.1	OG0000785
Bimnz	484389	Pleosporales	I	Effector	NLP1.2	OG0000785
Bimnz	524705	Pleosporales	I	Effector	NLP1.1	OG0000785
Botdo	286586	Botryosphaeriales	I	Non-effector	NLP1.3	OG0000785
Botdo	289801	Botryosphaeriales	I	Non-effector	NLP1.1	OG0000785
Botdo	289813	Botryosphaeriales	I	Non-effector	NLP1.3	OG0000785
Botdo	290603	Botryosphaeriales	I	Non-effector	NLP1.3	OG0000785
Botdo	294312	Botryosphaeriales	I	Non-effector	NLP1.1	OG0000785
Botdo	294542	Botryosphaeriales	I	Non-effector	NLP1.3	OG0000785
Bysci	416094	Pleosporales	I	Effector	NLP1.1	OG0000785
Bysci	495000	Pleosporales	I	Non-effector	NLP1.2	OG0000785
Bysci	585764	Pleosporales	II	Effector	NLP2	OG0010037
Cerzm	61174	Capnodiales	I	Effector	NLP1.1	OG0000785
Clael	447446	Pleosporales	I	Effector	NLP1.1	OG0000785
Clafu	194232	Capnodiales	I	Non-effector	NLP1.1	OG0000785
Coche	1032627	Pleosporales	I	Effector	NLP1.1	OG0000785
Coche	1171698	Pleosporales	II	Effector	NLP2	OG0010037
Cocsa	33268	Pleosporales	II	Non-effector	NLP2	OG0010037
Cocsa	89541	Pleosporales	I	Effector	NLP1.1	OG0000785
Cocsa	135644	Pleosporales	I	Effector	NLP1.3	OG0000785
Corca	497672	Pleosporales	I	Effector	NLP1.3	OG0000785
Corca	537044	Pleosporales	I	Effector	NLP1.2	OG0000785
Corca	601720	Pleosporales	I	Effector	NLP1.1	OG0000785
Corol[1]	g8735	Pleosporales	I	Non-effector	NLP1.1	OG0000785
Didex	91007	Pleosporales	I	Effector	NLP1.1	OG0000785
Didex	91324	Pleosporales	II	Effector	NLP2	OG0010037
Didex	418130	Pleosporales	I	Non-effector	NLP1.3	OG0000785
Dipse	4716	Botryosphaeriales	I	Effector	NLP1.3	OG0000785
Dipse	5251	Botryosphaeriales	I	Effector	NLP1.1	OG0000785
Dipse	9320	Botryosphaeriales	I	Effector	NLP1.3	OG0000785
Dipse	9327	Botryosphaeriales	I	Effector	NLP1.1	OG0000785
Hyspu	116085	Hysteriales	I	Non-effector	NLP1.1	OG0000785
Karrh	368783	Pleosporales	I	Effector	NLP1.1	OG0000785
Lenfl	393400	Pleosporales	I	Non-effector	NLP1.1	OG0000785
Lepma	12469	Pleosporales	I	Non-effector	NLP1.3	OG0000785
Lepma	11090	Pleosporales	I	Effector	NLP1.1	OG0000785
Macph	2384	Botryosphaeriales	I	Effector	NLP1.3	OG0000785
Macph	6483	Botryosphaeriales	I	Effector	NLP1.1	OG0000785
Macph	10738	Botryosphaeriales	I	Effector	NLP1.3	OG0000785
Macph	12781	Botryosphaeriales	I	Non-effector	NLP1.3	OG0000785
Maseb	418354	Pleosporales	I	Effector	NLP1.1	OG0000785
Mycfi	78817	Capnodiales	II	Effector	NLP2	OG0010037
Mycfi	83654	Capnodiales	I	Non-effector	NLP1.1	OG0000785
Neopa	727	Botryosphaeriales	I	Effector	NLP1.1	OG0000785
Neopa	928	Botryosphaeriales	I	Effector	NLP1.3	OG0000785
Neopa	2549	Botryosphaeriales	I	Effector	NLP1.3	OG0000785
Neopa	6217	Botryosphaeriales	I	Non-effector	NLP1.3	OG0000785
Neopa	6314	Botryosphaeriales	I	Effector	NLP1.2	OG0000785
Neopa	7612	Botryosphaeriales	I	Effector	NLP1.1	OG0000785
Parsp	876375	Pleosporales	I	Effector	NLP1	OG0000785
Parsp	1126479	Pleosporales	I	Effector	NLP1.1	OG0000785
Photr	488114	Pleosporales	I	Effector	NLP1.1	OG0000785
Rhyru	113036	Hysteriales	I	Non-effector	NLP1.1	OG0000785
Sepmu	74880	Capnodiales	I	Effector	NLP1.1	OG0000785
Seppo	56247	Capnodiales	I	Effector	NLP1.1	OG0000785
Settu	41216	Pleosporales	I	Effector	NLP1.1	OG0000785
Stano	3364	Pleosporales	I	Effector	NLP1.1	OG0000785
Stano	7008	Pleosporales	II	Effector	NLP2	OG0010037
Trepe	429691	Pleosporales	I	Non-effector	NLP1.1	OG0000785

[1] Gene ID according to gene prediction by Augustus (Chapter 1).

Narrow sense phylogeny of type I and type II NLPs

In the subsequent analyses, we reconstruct independent phylogenetic analyses for each of the three subclades of type I NLP we had uncovered in the broad sense phylogeny. We also build an independent phylogeny for type II NLPs (Figure 3, Table 2). Hereafter, the three narrow sense phylogenies of type I NLPs are referred to as 'NLP1.1', 'NLP1.2', and 'NLP1.3', respectively.

By far, NLP1.1 was the most member-rich phylogeny, comprising 33 type I NLPs (24 effectors, 76%) from 29 species of Dothideomycetes and one NLP from *A. nidulans* (Figure 3a). For those 14 species that contained a single NLP; without exception, those NLPs were present in NLP1.1. Among the 32 NLPs of Dothideomycetes, there were a total of 23 NLP effectors. The richness towards NLP effectors was higher among Pleosporales, in which 15 of 18 were predicted as effector proteins.

From a polytomy (Figure 3a), clade 1 splits into the subclade of Capnodiales and the subclade of Botryospaeriales. Next, there was clade 2 that contains Botryospaeriales, exclusively. Finally, clade 3 was the largest of the three clades, and contains *Cercospora zeae-maydis* (Capnodiales), followed by *Rhytidhysteron rufulum* and *Hysterium pulicare* (Hysteriales), and finally the large sub-clade of Pleosporales. Overall, NLP1.1 paralleled quite well the molecular phylogenetic framework of Dothideomycetes, especially across Pleosporales. Botryosphaeriales retained a major gene duplication of NLP1.1, with sets of corresponding paralogs in distinct genera being retained over time (clades 1 and 2). This major gene duplication is absent from either Capnodiales or Pleosporales.

The type I NLP that are members of the NLP1.2 were of rare occurrence; this phylogeny contained only four NLPs (three effectors) that were recovered from four species of Dothideomycetes (Figure 3b).

The remaining 17 type I NLPs from nine species of Dothideomycetes were analyzed in NLP1.3; about half of the members (9 of 17) were non-effectors NLP (Figure 3c, Table 2). Once again, this phylogeny revealed two additional duplication events, one of them was shared only among Botryosphaeriales. The other duplication gave rise to a sub-clade that has a sister relationship to Pleosporales. Although not as member-rich as NLP1.1, NLP1.3 also matched well the molecular phylogenetic framework of Dothideomycetes.

The type II NLPs contained was also of rare occurrence (6 of 33 species) amongst the Dothideomycetes, but five species were Pleosporales (Figure 3d). Despite the rarity, the topology of type II NLPs still resembled the molecular phylogenetic framework of Dothideomycetes.

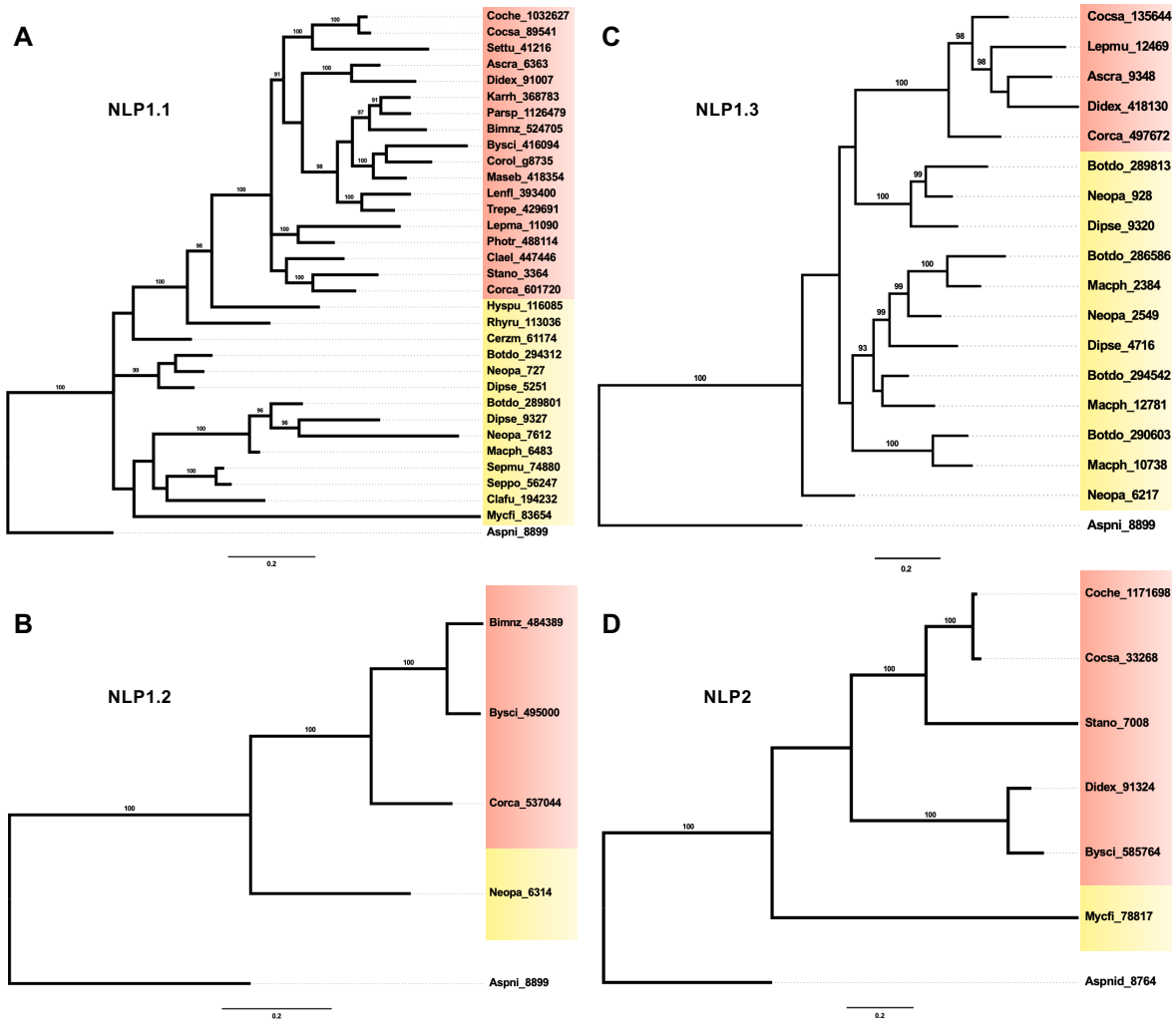


Figure 3. Phylogenetic relationships among members of each subclade of Necrosis- and Ethylene-inducing Like proteins (NLP) in Dothideomycetes. Each phylogeny represents a subclade in Figure 2 (and respective dataset) as followed: **A.** Members of NLP1.1 (dataset comprised proteins that were 295 amino acids long); **B.** Members of NLP1.2 (comprised proteins that were 254 amino acids long); **C.** Members of NLP1.3 (comprised proteins that were 274 amino acids long); **D.** Members of NLP2 (comprised proteins that were 422 amino acids long). For each phylogeny, one NLP from *Aspergillus nidulans* (Eurotiomycetes) was used as outgroup, as indicated. Clades are color-coded according to their placements within Dothideomycetes

(orange, Pleosporales; yellow, non-Pleosporales. Branch lengths are drawn to scale; nodal support values are given as posterior probabilities (when ≥ 90%) above the branches. Scale bar corresponds to the expected number of substitutions per site.

Cytotoxic NLPs within a phylogenetic context

Our survey recovered 27 functional analyses of NLPs (Table 3). There were 11 instances in which the authors attributed to the NLP under investigation a strong activity, all of the proteins were type I NLPs. Among those 11 proteins with strong cytotoxic activity, there were seven NLP1.1, three NLP1.2, and one NLP1.3. There were nine instances in which a type 1 NLP showed weak/absent activity; those instances took place in species with multiple copies of the type I NLP and a strong activity already had been attributed to one of the NLP1 paralogs. Finally, all six functional analyses of type II NLPs (NLP2) and a single analysis of type III NLP pointed to weak/absent activity. In *Neofusicoccum parvum* (Botryosphaeriales), the NLP1.3 paralogs Neopa_928 and Neopa_6217 had been both removed from functional analysis experiment (Pour et al., 2020). While the overexpression of Neopa_928 as a recombinant protein was not possible; Neopa_6217 coded for a truncated protein that lacked signal peptide. Meanwhile, our analyses predicted that both Neopa_928 was an NLP effector and Neopa_6217 was a non-effector NLP.

Table 3. The cytotoxicity of NLPs found in species of the phylum Ascomycota.

Reference	Class/Order/Species	NLP name	Cytotoxicity to dicots	Gene ID (this work)	NLP context (this work)
Cobos et al., (2019)	Dothideomycetes/Botryosphaeriales/*Diplodia seriata*	DserNEP1	Strong	Dipse_5251	NLP1.1
Cobos et al., (2019)	Dothideomycetes/Botryosphaeriales/*Diplodia seriata*	DserNEP2	Weak	Dipse_9327	NLP1.1
Pour et al., (2020)	Dothideomycetes/Botryosphaeriales/*Neofusicoccum parvum*	NprvNep1	Strong	Neopa_727	NLP1.1
Pour et al., (2020)	Dothideomycetes/Botryosphaeriales/*Neofusicoccum parvum*	NprvNep2	Strong	Neopa_7612	NLP1.1
Pour et al., (2020)	Dothideomycetes/Botryosphaeriales/*Neofusicoccum parvum*	NprvNep3	Weak	Neopa_6314	NLP1.2
Pour et al., (2020)	Dothideomycetes/Botryosphaeriales/*Neofusicoccum parvum*	NprvNep4	Weak	Neopa_2549	NLP1.3
Pour et al. (2020)	Dothideomycetes/Botryosphaeriales/*Neofusicoccum parvum*	NprvNep5	-	Neopa_928[2]	NLP1.3
Pour et al. (2020)	Dothideomycetes/Botryosphaeriales/*Neofusicoccum parvum*	NprvNep6	-	Neopa_6217[2]	NLP1.3
Motteram et al., (2009)	Dothideomycetes/Capnodiales/*Zymoseptoria tritici*	MgNLP	Strong	-	NLP1.1[3]
Levin et al., (2019)	Eurotiomycetes/Eurotiales/*Penicillium expansum*	PeNLP1	Strong	-	NLP1.1[3]
Levin et al., (2019)	Eurotiomycetes/Eurotiales/*Penicillium expansum*	PeNLP2[1]	Weak	-	-
Santhanam et al., (2013)	Sordariomycetes/Hypocreales/*Verticillium dahliae*	NLP1	Strong	-	NLP1.2[3]
Santhanam et al., (2013)	Sordariomycetes/Hypocreales/*Verticillium dahliae*	NLP2	Weak	-	NLP1.1[3]
Santhanam et al., (2013)	Sordariomycetes/Hypocreales/*Verticillium dahliae*	NLP3	Absent	-	NLP1.2[3]
Santhanam et al., (2013)	Sordariomycetes/Hypocreales/*Verticillium dahliae*	NLP4	Absent	-	NLP2[3]
Santhanam et al., (2013)	Sordariomycetes/Hypocreales/*Verticillium dahliae*	NLP5	Absent	-	NLP2[3]
Santhanam et al., (2013)	Sordariomycetes/Hypocreales/*Verticillium dahliae*	NLP6	Absent	-	NLP1.2[3]
Santhanam et al., (2013)	Sordariomycetes/Hypocreales/*Verticillium dahliae*	NLP7	Absent	-	NLP2[3]
Santhanam et al., (2013)	Sordariomycetes/Hypocreales/*Verticillium dahliae*	NLP9	Absent	-	NLP2[3]
Fang et al., (2017)	Sordariomycetes/Magnaporthales/*Magnaporthe oryzae*	MoNLP1	Strong	-	NLP1.1[3]
Fang et al., (2017)	Sordariomycetes/Magnaporthales/*Magnaporthe oryzae*	MoNLP2	Weak	-	NLP2[3]
Fang et al., (2017)	Sordariomycetes/Magnaporthales/*Magnaporthe oryzae*	MoNLP3	Absent	-	NLP2[3]
Fang et al., (2017)	Sordariomycetes/Magnaporthales/*Magnaporthe oryzae*	MoNLP4	Strong	-	NLP1.3[3]
Bashi et al., (2010)	Leotiomycetes/Helotiales/*Sclerotinia sclerotiorum*	SsNLP1	Weak	-	NLP1.2[3]
Bashi et al., (2010)	Leotiomycetes/Helotiales/*Sclerotinia sclerotiorum*	SsNLP2	Strong	-	NLP1.1[3]
Arenas et al., (2010)	Leotiomycetes/Helotiales/*Botrytis cinerea*	BcNEP1	Strong	-	NLP1.2[3]
Arenas et al., (2010)	Leotiomycetes/Helotiales/*Botrytis cinerea*	BcNEP2	Weak	-	NLP1.1[3]
Staats et al., (2007)	Leotiomycetes/Helotiales/*Botrytis elliptica*	BeNEP1	Strong	-	NLP1.2[3]
Staats et al., (2007)	Leotiomycetes/Helotiales/*Botrytis elliptica*	BeNEP2	Weak	-	NLP1.1[3]

[1] PeNLP2 is type III NLP.

[2] Neopa_928 and Neopa_6217 were predicted to encode an NLP effector and a non-effector NLP, respectively (Table 2)

[3] Proteins that were individually aligned against our sequence database, but data not included in the phylogenetic treatment (Figure 2).

DISCUSSION

The NLP superfamily

The current type-based classification of NLPs (types I, II, and III) relies on the number of conserved cysteine residues that form disulfide bridges along the N-terminal half of the proteins (Gijzen and Nürnberger, 2006; Oome and Van Den Ackerveken, 2014). Previously, a broad phylogenetic analysis of NLPs from bacteria, fungi and oomycetes provided strong support to postulate that NLPs of a given type shared a common ancestor (Seidl and Van den Ackerveken, 2019). Using phylogenetic data from the Dothideomycetes, we demonstrated herein that the phylogenetic distribution of NLPs followed closely the phylogenomic relationships of the genera in which the NLPs were found.

Therefore, the strong concordance between phylogenomics and NLP phylogeny support a phylogeny-based classification of the NLP superfamily, with at least two subordinate families: hereafter we will refer to type I NLPs collectively as the 'NLP1 family', and to type II NLPs as the 'NLP2 family'. Furthermore, the phylogenetic relationships within the member-rich NLP1 family of Dothideomycetes supported three further natural subdivisions: hereafter referred to as NLP1.1, NLP1.2, and NLP1.3 subfamilies. Likely, type III NLPs also merit their own phylogenetic-based category; however, several aspects of their evolutionary history are still unclear, such as their low sequence conservation and likely association with HGT (Seidl and Van den Ackerveken, 2019); thus, we refrained from making further inferences about them.

Biased patterns drove NLP1 gene evolution

The repositories of NLPs in the genomes of extant Dothideomycetes seem to be the product of dynamic events that took place along their evolutionary trajectories. The NLP1 and NLP2 families clearly underwent distinct pathways during the evolution of the Dothideomycetes. Gene loss seems to have played a major role in shaping the evolutionary history of the NLP1.2 subfamily and the NLP2 family, with some few extant homologs distributed all over the phylogeny of that class (Figure 3d). Meanwhile, the NLP1 family was of ubiquitous occurrence and underwent rounds of gene duplication followed by descent with modification.

The signatures of those past lineage-specific gene duplication and gene retention are visible in the phylogenies of subfamilies NLP1.1 and NLP1.3, especially

in the order Botryosphaeriales, which showed above-average number of the NLP1 family (Figures 3a and 3c). The otherwise redundancy of copies may actually represent a strategy of functional diversification to overcome host-related defense mechanisms, with distinct NLPs playing a functional role in a time-dependent manner to allow successful colonization of plant tissues.

Among Dothideomycetes, NLP superfamily size may be associated with fungal lifestyle, as necrotrophic plant pathogens detained the highest number of NLP copies. In these pathogens, cell death-inducing proteins that act at the plant apoplast level, such are NLPs, are essential for host colonization (Cobos et al., 2019; Li et al., 2020; Pour et al., 2020). Moreover, the role of larger NLP family sizes in such necrotrophic plant pathogens can go beyond differential cytotoxicity and detain levels of functional diversification at different life stages (Santhanam et al., 2013; Seidl et al., 2015). Finally, broader host ranges, such as described by *N. parvum* (Pour et al., 2020) and *C. cassiicola* (Dixon et al., 2009; Lopez et al., 2018), could also contribute to unbalance the family size, favoring polyphagous plant pathogenetic species over host specific ones (Baroncelli et al., 2016; Kleemann et al., 2012).

Phylogenetic framework for NLP cytotoxicity

Cytotoxic and non-cytotoxic NLPs can coexist in the same species (Dong et al., 2012) and they are usually differed by the presence of a secretion signal in the cytotoxic proteins (Levin et al., 2019; Oome et al., 2014). Here, we chose to investigate the likely function of NLPs based on the presence (effectors) or absence (non-effectors) of a predicted signal peptide together with a survey on functional activities of NLPs. The distribution of non-effector NLPs across the phylogeny showed that the classification of NLPs as effectors cannot be taken as monophyletic clades, suggesting independent losses for the secretion signal along evolution, possibility to escape from host recognition.

The vast majority of NLP effectors in the cytotoxic-active member rich subfamily NLP1.1 suggests for preferential maintenance of these proteins due to their functional activity. *In vitro* deletion of a member of the NLP1 family led to the reduction of virulence in the postharvest pathogen *Penicillium expansum* (Eurotiomycetes); meanwhile the virulence remained unchanged when a type III NLP was the deleted copy (Levin et al., 2019). In *N. parvum*, a NLP1.1 member was strongly cytotoxic to

both tomato plants and mammalian cells, while its paralogous from others NLP1 subfamilies showed minor activity (Pour et al., 2020).

Proteomic analyses of the secretome found only two of the four NLPs of *D. seriata*, both of them belonged to the NLP1.1 subfamily but showed distinct levels of cytotoxicity to grapevine leaves (Cobos et al., 2019, 2010). The only NLP from *Zymoseptoria tritici*, which induced defense responses and cell death in dicots but not in monocots (Motteram et al., 2009), shared homology to NLP1.1. Finally, when taken altogether, those previous reports support our hypothesis that cytotoxic activity of NLPs is associated with preferential gene retention, especially of members of the NLP1 family in Dothideomycetes.

Acknowledgement

This work was supported by The Minas Gerais State Foundation of Research Aid – FAPEMIG (grant number APQ-00150-17) and by The National Council of Scientific and Technological Development – CNPq (fellowship number PQ 302336/2019-2) to LOO. TCSD received student fellowships from the CAPES Foundation (PROEX – 0487 No. 1684083) from Mar/2017 to Jul/2018 and CNPq (GM/GD 142400/2018-1) from Aug/2018 to Apr/2021.

Conflict of interest

The authors have declared that no competing interests exist.

REFERENCES

Amsellem, Z., Cohen, B.A., Gressel, J., 2002. Engineering hypervirulence in a mycoherbicidal fungus for efficient weed control. Nat. Biotechnol. 20, 1035–1039.

Arenas, Y.C., Kalkman, E.R.I.C., Schouten, A., Dieho, M., Vredenbregt, P., Uwumukiza, B., Osés Ruiz, M., van Kan, J.A.L., 2010. Functional analysis and mode of action of phytotoxic Nep1-like proteins of *Botrytis cinerea*. Physiol. Mol. Plant Pathol. 74, 376–386.

Bailey, B.A., 1995. Purification of a protein from culture filtrates of *Fusarium oxysporum* that induces ethylene and necrosis in leaves of *Erythroxylum coca*. Phytopathology 85, 1035–1039.

Baroncelli, R., Amby, D.B., Zapparata, A., Sarrocco, S., Vannacci, G., Floch, G. Le, Harrison, R.J., Holub, E., Sukno, S.A., Sreenivasaprasad, S., Thon, M.R., 2016. Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. BMC Genomics 17, 555.

Bashi, Z.D., Hegedus, D.D., Buchwaldt, L., Rimmer, S.R., Borhan, M.H., 2010. Expression and regulation of *Sclerotinia sclerotiorum* necrosis and ethylene-inducing peptides (NEPs). Mol. Plant Pathol. 11, 43–53.

Blanco-Ulate, B., Rolshausen, P., Cantu, D., 2013. Draft genome sequence of *Neofusicoccum parvum* isolate UCR-NP2, a fungal vascular pathogen associated with grapevine cankers. Genome Announc. 1.

Böhm, H., Albert, I., Oome, S., Raaymakers, T.M., Van den Ackerveken, G., Nürnberger, T., 2014. A conserved peptide pattern from a widespread microbial virulence factor triggers pattern-induced immunity in *Arabidopsis.* PLoS Pathog. 10.

Buchfink, B., Xie, C., Huson, D.H., 2014. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60.

Cabral, A., Oome, S., Sander, N., Küfner, I., Nürnberger, T., Van Den Ackerveken, G., 2012. Nontoxic Nep1-like proteins of the downy mildew pathogen *Hyaloperonospora arabidopsidis*: Repression of necrosis-inducing activity by a surface-exposed region. Mol. Plant-Microbe Interact. 25, 697–708.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.

Cobos, R., Barreiro, C., Mateos, R.M., Coque, J.J.R., 2010. Cytoplasmic- and extracellular-proteome analysis of *Diplodia seriata*: A phytopathogenic fungus involved in grapevine decline. Proteome Sci. 8, 1–16.

Cobos, R., Calvo-Peña, C., Álvarez-Pérez, J.M., Ibáñez, A., Diez-Galán, A., González-García, S., García-Angulo, P., Acebes, J.L., Coque, J.J.R., 2019. Necrotic and cytolytic activity on grapevine leaves produced by Nep1-like proteins of *Diplodia seriata*. Front. Plant Sci. 10, 1–13.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164–1165.

Dixon, L.J., Schlub, R.L., Pernezny, K., Datnoff, L.E., 2009. Host specialization and phylogenetic diversity of *Corynespora cassiicola*. Phytopathology 99, 1015–1027.

de Wit, P.J.G.M., van der Burgt, A., Ökmen, B., Stergiopoulos, I., Abd-Elsalam, K.A., Aerts, A.L., Bahkali, A.H., Beenen, H.G., Chettri, P., Cox, M.P., Datema, E., de Vries, R.P., Dhillon, B., Ganley, A.R., Griffiths, S.A., Guo, Y., Hamelin, R.C., Henrissat, B., Kabir, M.S., Jashni, M.K., Kema, G., Klaubauf, S., Lapidus, A., Levasseur, A., Lindquist, E., Mehrabi, R., Ohm, R.A., Owen, T.J., Salamov, A., Schwelm, A., Schijlen, E., Sun, H., van den Burg, H.A., van Ham, R.C.H.J., Zhang, S., Goodwin, S.B., Grigoriev, I. V, Collemare, J., Bradshaw, R.E., 2012. The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaptation to different hosts and lifestyles but also signatures of common ancestry. PLOS Genet. 8, 1–22.

Dong, S., Kong, G., Qutob, D., Yu, X., Tang, J., Kang, J., Dai, T., Wang, H., Gijzen, M., Wang, Y., 2012. The NLP toxin family in *Phytophthora sojae* includes rapidly evolving groups that lack necrosis-inducing activity. Mol. Plant-Microbe Interact. 25, 896–909.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., Finn, R.D., 2019. The Pfam protein families database in 2019. Nucleic Acids Res. 47, D427–D432.

Emanuelsson, O., Brunak, S., von Heijne, G., Nielsen, H., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. 2, 953–971.

Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 1–14.

Fang, Y.L., Peng, Y.L., Fan, J., 2017. The Nep1-like protein family of *Magnaporthe oryzae* is dispensable for the infection of rice plants. Sci. Rep. 7, 1–10.

Fellbrich, G., Romanski, A., Varet, A., Blume, B., Brunner, F., Engelhardt, S., Felix, G., Kemmerling, B., Krzymowska, M., Nürnberger, T., 2002. NPP1, a *Phytophthora*-associated trigger of plant defense in parsley and *Arabidopsis*. Plant J. 32, 375–390.

Gijzen, M., Nürnberger, T., 2006. Nep1-like proteins from plant pathogens: Recruitment and diversification of the NPP1 domain across taxa. Phytochemistry 67, 1800–1807.

Grigoriev, I. V, Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., Shabalov, I., 2013. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. 42, D699.

Hane, J.K., Lowe, R.G.T., Solomon, P.S., Tan, K.C., Schoch, C.L., Spatafora, J.W., Crous, P.W., Kodira, C., Birren, B.W., Galagan, J.E., Torriani, S.F.F., McDonald, B.A., Oliver, R.P., 2007. Dothideomycete-plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. Plant Cell 19, 3347–3368.

Haridas, S., Albert, R., Binder, M., Bloem, J., LaButti, K., Salamov, A., Andreopoulos, B., Baker, S.E., Barry, K., Bills, G., Bluhm, B.H., Cannon, C., Castanera, R., Culley, D.E., Daum, C., Ezra, D., González, J.B., Henrissat, B., Kuo, A., Liang, C., Lipzen, A., Lutzoni, F., Magnuson, J., Mondo, S.J., Nolan, M., Ohm, R.A., Pangilinan, J., Park, H.J., Ramírez, L., Alfaro, M., Sun, H., Tritt, A., Yoshinaga, Y., Zwiers, L.H., Turgeon, B.G., Goodwin, S.B., Spatafora, J.W., Crous, P.W., Grigoriev, I. V., 2020. 101 Dothideomycetes genomes: A test case for predicting lifestyles and emergence of pathogens. Stud. Mycol. 96, 141–153.

Hyde, K.D., Jones, E.B.G., Liu, J.K., Ariyawansa, H., Boehm, E., Boonmee, S., Braun, U., Chomnunti, P., Crous, P.W., Dai, D.Q., Diederich, P., Dissanayake, A., Doilom, M., Doveri, F., Hongsanan, S., Jayawardena, R., Lawrey, J.D., Li, Y.M., Liu, Y.X., Lücking, R., Monkai, J., Muggia, L., Nelsen, M.P., Pang, K.L., Phookamsak, R., Senanayake, I.C., Shearer, C.A., Suetrong, S., Tanaka, K., Thambugala, K.M., Wijayawardene, N.N., Wikee, S., Wu, H.X., Zhang, Y., Aguirre-Hudson, B., Alias, S.A., Aptroot, A., Bahkali, A.H., Bezerra, J.L., Bhat, D.J., Camporesi, E., Chukeatirote, E., Gueidan, C., Hawksworth, D.L., Hirayama, K., De Hoog, S.,

Kang, J.C., Knudsen, K., Li, W.J., Li, X.H., Liu, Z.Y., Mapook, A., McKenzie, E.H.C., Miller, A.N., Mortimer, P.E., Phillips, A.J.L., Raja, H.A., Scheuer, C., Schumm, F., Taylor, J.E., Tian, Q., Tibpromma, S., Wanasinghe, D.N., Wang, Y., Xu, J.C., Yacharoen, S., Yan, J.Y., Zhang, M., 2013. Families of Dothideomycetes. Fungal Divers. 63, 1–313.

Isaza, R.E.A., Diaz-Trujillo, C., Dhillon, B., Aerts, A., Carlier, J., Crane, C.F., de Jong, T., de Vries, I., Dietrich, R., Farmer, A.D., Fortes Fereira, C., Garcia, S., Guzman, M., Hamelin, R.C., Lindquist, E.A., Mehrabi, R., Quiros, O., Schmutz, J., Shapiro, H., Reynolds, E., Scalliet, G., Souza Jr., M., Stergiopoulos, I., der Lee, T.A.J., De Wit, P.J.G.M., Zapater, M.-F., Zwiers, L.-H., Grigoriev, I. V, Goodwin, S.B., Kema, G.H.J., 2016. Combating a global threat to a clonal crop: Banana Black Sigatoka pathogen *Pseudocercospora fijiensis* (Synonym *Mycosphaerella fijiensis*) genomes reveal clues for disease control. PLOS Genet. 12, 1–36.

Islam, M.S., Haque, M.S., Islam, M.M., Emdad, E.M., Halim, A., Hossen, Q.M.M., Hossain, M.Z., Ahmed, B., Rahim, S., Rahman, M.S., Alam, M.M., Hou, S., Wan, X., Saito, J.A., Alam, M., 2012. Tools to kill: genome of one of the most destructive plant pathogenic fungi *Macrophomina phaseolina*. BMC Genomics 13, 493.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: Genome-scale protein function classification. Bioinformatics 30, 1236–1240.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Kirk, P., Cannon, P., Minter, D., Stalpers, J., 2008. Ainsworth and Bisby's dictionary of the Fungi, 10th ed. CAB International, Wallingford, UK.

Kleemann, J., Rincon-Rivera, L.J., Takahara, H., Neumann, U., van Themaat, E.V.L., van der Does, H.C., Hacquard, S., Stüber, K., Will, I., Schmalenbach, W., Schmelzer, E., O'Connell, R.J., 2012. Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. PLoS Pathog. 8.

Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E.L.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. 305, 567–580.

Levin, E., Raphael, G., Ma, J., Ballester, A.R., Feygenberg, O., Norelli, J., Aly, R., Gonzalez-Candelas, L., Wisniewski, M., Droby, S., 2019. Identification and functional analysis of nlp-encoding genes from the postharvest pathogen penicillium expansum. Microorganisms 7.

Li, Y., Han, Y., Qu, M., Chen, J., Chen, X., Geng, X., Wang, Z., Chen, S., 2020. Apoplastic cell death-inducing proteins of filamentous plant pathogens: Roles in plant-pathogen interactions. Front. Genet. 11.

Liu, J.K., Hyde, K.D., Jeewon, R., Phillips, A.J.L., Maharachchikumbura, S.S.N., Ryberg, M., Liu, Z.Y., Zhao, Q., 2017. Ranking higher taxa using divergence times: a case study in Dothideomycetes. Fungal Divers. 84, 75–99.

Lopez, D., Ribeiro, S., Label, P., Fumanal, B., Venisse, J.S., Kohler, A., Oliveira, R.R., Labutti, K., Lipzen, A., Lail, K., Bauer, D., Ohm, R.A., Barry, K.W., Spatafora, J., Grigoriev, I. V., Martin, F.M., Pujade-Renaud, V., 2018. Genome-wide analysis of *Corynespora cassiicola* leaf fall disease putative effectors. Front. Microbiol. 9, 1–21.

Marsberg, A., Kemler, M., Jami, F., Nagel, J.H., Postma-Smidt, A., Naidoo, S., Wingfield, M.J., Crous, P.W., Spatafora, J.W., Hesse, C.N., Robbertse, B., Slippers, B., 2017. Botryosphaeria dothidea: a latent pathogen of global importance to woody plant health. Mol. Plant Pathol. 18, 477–488.

Morales-Cruz, A., Amrine, K.C.H., Blanco-Ulate, B., Lawrence, D.P., Travadon, R., Rolshausen, P.E., Baumgartner, K., Cantu, D., 2015. Distinctive expansion of gene families associated with plant cell wall degradation, secondary metabolism, and nutrient uptake in the genomes of grapevine trunk pathogens. BMC Genomics 16, 469.

Motteram, J., Küfner, I., Deller, S., Brunner, F., Hammond-Kosack, K.E., Nürnberger, T., Rudd, J.J., 2009. Molecular characterization and functional analysis of MgNLP, the sole NPP1 domain-containing protein, from the fungal wheat leaf pathogen *Mycosphaerella graminicola*. Mol. Plant-Microbe Interact. 22, 790–799.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.

Ohm, R.A., Feau, N., Henrissat, B., Schoch, C.L., Horwitz, B.A., Barry, K.W., Condon, B.J., Copeland, A.C., Dhillon, B., Glaser, F., Hesse, C.N., Kosti, I., LaButti, K., Lindquist, E.A., Lucas, S., Salamov, A.A., Bradshaw, R.E., Ciuffetti, L., Hamelin, R.C., Kema, G.H.J., Lawrence, C., Scott, J.A., Spatafora, J.W., Turgeon, B.G., de Wit, P.J.G.M., Zhong, S., Goodwin, S.B., Grigoriev, I. V., 2012. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. PLoS Pathog. 8.

Oome, S., Raaymakers, T.M., Cabral, A., Samwel, S., Böhm, H., Albert, I., Nürnberger, T., Van Den Ackerveken, G., 2014. Nep1-like proteins from three kingdoms of life act as a microbe-associated molecular pattern in *Arabidopsis.* Proc. Natl. Acad. Sci. U. S. A. 111, 16955–16960.

Oome, S., Van Den Ackerveken, G., 2014. Comparative and functional analysis of the widely occurring family of Nep1-like proteins. Mol. Plant-Microbe Interact. 27, 1081–1094.

Ottmann, C., Luberacki, B., Küfner, I., Koch, W., Brunner, F., Weyand, M., Mattinen, L., Pirhonen, M., Anderluh, G., Seitz, H.U., Nürnberger, T., Oecking, C., 2009. A common toxin fold mediates microbial attack and plant defense. Proc. Natl. Acad. Sci. U. S. A. 106, 10359–10364.

Pemberton, C.L., Salmond, G.P.C., 2004. The Nep1-like proteins - A growing family of microbial elicitors of plant necrosis. Mol. Plant Pathol. 5, 353–359.

Peter, M., Kohler, A., Ohm, R.A., Kuo, A., Krützmann, J., Morin, E., Arend, M., Barry, K.W., Binder, M., Choi, C., Clum, A., Copeland, A., Grisel, N., Haridas, S., Kipfer, T., LaButti, K., Lindquist, E., Lipzen, A., Maire, R., Meier, B., Mihaltcheva, S., Molinier, V., Murat, C., Pöggeler, S., Quandt, C.A., Sperisen, C., Tritt, A., Tisserant, E., Crous, P.W., Henrissat, B., Nehls, U., Egli, S., Spatafora, J.W., Grigoriev, I. V, Martin, F.M., 2016. Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus *Cenococcum geophilum*. Nat. Commun. 7, 12662.

Petersen, T.N., Brunak, S., Von Heijne, G., Nielsen, H., 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.

Pour, F.N., Cobos, R., Coque, J.J.R., Serôdio, J., Alves, A., Félix, C., Ferreira, V., Esteves, A.C., Duarte, A.S., 2020. Toxicity of recombinant Necrosis and Ethylene-Inducing Proteins (NLPs) from *Neofusicoccum parvum*. Toxins. 12, 1–17.

Qutob, D., Kemmerling, B., Brunner, F., Küfner, I., Engelhardt, S., Gust, A.A., Luberacki, B., Seitz, H.U., Stahl, D., Rauhut, T., Glawischnig, E., Schween, G., Lacombe, B., Watanabe, N., Lam, E., Schlichting, R., Scheel, D., Nau, K., Dodt, G., Hubert, D., Gijzen, M., Nürnberger, T., 2006. Phytotoxicity and innate immune responses induced by Nep1-like proteins. Plant Cell 18, 3721–3744.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67, 901–904.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Rouxel, T., Grandaubert, J., Hane, J.K., Hoede, C., Van De Wouw, A.P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., Cozijnsen, A.J., Ciuffetti, L.M., Degrave, A., Dilmaghani, A., Duret, L., Fudal, I., Goodwin, S.B., Gout, L., Glaser, N., Linglin, J., Kema, G.H.J., Lapalu, N., Lawrence, C.B., May, K., Meyer, M., Ollivier, B., Poulain, J., Schoch, C.L., Simon, A., Spatafora, J.W., Stachowiak, A., Turgeon, B.G., Tyler, B.M., Vincent, D., Weissenbach, J., Amselem, J., Quesneville, H., Oliver, R.P., Wincker, P., Balesdent, M.H., Howlett, B.J., 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by repeat-induced point mutations. Nat. Commun. 2.

Santhanam, P., Van Esse, H.P., Albert, I., Faino, L., Nürnberger, T., Thomma, B.P.H.J., 2013. Evidence for functional diversification within a fungal Nep1-like protein family. Mol. Plant-Microbe Interact. 26, 278–286.

Schoch, C.L., Crous, P.W., Groenewald, J.Z., Boehm, E.W.A., Burgess, T.I., de Gruyter, J., de Hoog, G.S., Dixon, L.J., Grube, M., Gueidan, C., Harada, Y., Hatakeyama, S., Hirayama, K., Hosoya, T., Huhndorf, S.M., Hyde, K.D., Jones, E.B.G., Kohlmeyer, J., Kruys, Å., Li, Y.M., Lücking, R., Lumbsch, H.T., Marvanová, L., Mbatchou, J.S., McVay, A.H., Miller, A.N., Mugambi, G.K., Muggia, L., Nelsen, M.P., Nelson, P., Owensby, C.A., Phillips, A.J.L., Phongpaichit, S., Pointing, S.B., Pujade-Renaud, V., Raja, H.A., Plata, E.R., Robbertse, B., Ruibal, C., Sakayaroj, J., Sano, T., Selbmann, L., Shearer, C.A., Shirouzu, T., Slippers, B., Suetrong, S., Tanaka, K., Volkmann-Kohlmeyer, B., Wingfield, M.J., Wood, A.R., Woudenberg,

J.H.C., Yonezawa, H., Zhang, Y., Spatafora, J.W., 2009. A class-wide phylogenetic assessment of Dothideomycetes. Stud. Mycol. 64, 1–15

Seidl, M.F., Faino, L., Shi-Kunne, X., Van Den Berg, G.C.M., Bolton, M.D., Thomma, B.P.H.J., 2015. The genome of the saprophytic fungus *Verticillium tricorpus* reveals a complex effector repertoire resembling that of its pathogenic relatives. Mol. Plant-Microbe Interact. 28, 362–373.

Seidl, M.F., Van den Ackerveken, G., 2019. Activity and phylogenetics of the broadly occurring family of microbial Nep1-Like Proteins. Annu. Rev. Phytopathol. 57, 367–386.

Staats, M., van Baarlen, P., Schouten, A., van Kan, J.A.L., Bakker, F.T., 2007. Positive selection in phytotoxic protein-encoding genes of *Botrytis* species. Fungal Genet. Biol. 44, 52–63.

Verma, S., Gazara, R.K., Nizam, S., Parween, S., Chattopadhyay, D., Verma, P.K., 2016. Draft genome sequencing and secretome analysis of fungal phytopathogen *Ascochyta rabiei* provides insight into the necrotrophic effector repertoire. Sci. Rep. 6, 1–14.

Zeiner, C.A., Purvine, S.O., Zink, E.M., Paša-Tolić, L., Chaput, D.L., Haridas, S., Wu, S., LaButti, K., Grigoriev, I. V., Henrissat, B., Santelli, C.M., Hansel, C.M., 2016. Comparative analysis of secretome profiles of manganese(II)-oxidizing Ascomycete fungi. PLoS One 11.

Chapter 3

Evolutionary history and expression analysis of the Necrosis- and Ethylene-inducing peptide 1-like protein (NLP) gene family in *Corynespora cassiicola*

Evolutionary history and expression analysis of the Necrosis- and Ethylene-inducing peptide 1-like protein (NLP) gene family in *Corynespora cassiicola*

Thaís Carolina da Silva Dal'Sasso[1], Vinícius Delgado da Rocha[1], Hugo Vianna Silva Rody[2], Maximiller Dal-Bianco Lamas Costa[1], Luiz Orlando de Oliveira[1]

[1]Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Viçosa, Viçosa, Brazil
[2]Departamento de Genética, Universidade de São Paulo/Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, Brazil

Corresponding author: Luiz Orlando de Oliveira
Email: lorlando@ufv.br

ABSTRACT

Effectors are proteins secreted by plant-associated pathogens to modify the structure and function of the host cell and facilitate the infection process. As effectors, the Necrosis- and Ethylene-inducing peptide 1-like proteins (NLPs) are involded in the early phases of plant infection and may trigger host immune responses. *Corynespora cassiicola* is a polyphagous plant-pathogen that causes target spot on many agriculturally important crops. We carried out comparative genomic analyses on 44 isolates of *C. cassiicola* and used an array of tools to explore the evolutionary history of their NLPs. We accessed the expression profiles of the NLP-encoding genes of *C. cassiicola* in the early hours of a pathogen-soybean interaction. Three NLP effector members (Cc_NLP1.1, Cc_NLP1.2, and Cc_NLP1.3) were maintained in the genomes of all isolates tested and a rare non-effector NLP member (Cc_NLP1.4) was found in three isolates that had been originally obtained from soybean. NLP effector members were under different selective constraints: Cc_NLP1.1 was under stronger selective pressure, while Cc_NLP1.2 was under a more relaxed constraint. Meanwhile, Cc_NLP1.3 was likely evolving under positive and balancing selection. Despite highly divergent, the NLP effectors maintain conserved the residues necessary to trigger plant immune responses. Gene expression profiles suggest that only Cc_NLP1.1 played a significant role in the early hours of host colonization.

Keywords: Effector genes, molecular evolution, selection, gene genealogy

INTRODUCTION

Effectors are molecules secreted by plant-associated microorganisms to modify the structure and function of the host cell, and by doing so, they facilitate the infection process (Lo Presti et al., 2015). In turn, plants have evolved receptors located at the cell surface and intracellular receptors, both of which able to detect conserved microbe-associated molecular patters (MAMPs) (Franceschetti et al., 2017). Adaptation to new hosts and evasion of pathogen recognition may require rapid changes in the effector repertoire, leading to rapid evolution of these genes (Fouché et al., 2018; Lo Presti et al., 2015; Sánchez-Vallet et al., 2018). The discovery of effector proteins that share conserved domains has started a new direction in plant-pathogen interaction studies. The identification of domains makes possible to predict effectors based on pathogen genome annotation, to determine the structure and to predict the function of these proteins in the host cell (Jones et al., 2018; Sonah et al., 2016). These conserved regions generally characterize proteins belonging to the same gene family and can be share even by phylogenetically distant organisms (Franceschetti et al., 2017; Liu et al., 2019).

Necrosis- and Ethylene-inducing peptide 1-like proteins (NLPs) constitute a superfamily found in a wide range of plant-associated microorganisms, including fungi, oomycetes, and bacteria (Gijzen and Nürnberger, 2006; Oome and Van Den Ackerveken, 2014; Seidl and Van den Ackerveken, 2019). They are usually considered effectors due to their cytotoxicity in eudicot plants that leads to the induction of cell death by necrosis and plant immune responses (Bae et al., 2006; Fellbrich et al., 2002; Lenarčič et al., 2017; Qutob et al., 2006). Although noncytolytic NLPs cannot permeabilize the plant membrane, they retain the capability of triggering plant immune responses (Cabral et al., 2012; Dong et al., 2012; Ottmann et al., 2009; Qutob et al., 2006).

NLPs are characterized by the presence of an NPP1 domain (Pfam PF05630) (Fellbrich et al., 2002). In the central part, the domain has a highly conserved amino acid motif (GHRHDWE) that is involved in cation binding (Böhm et al., 2014; Gijzen and Nürnberger, 2006; Oome et al., 2014; Ottmann et al., 2009). Based on the number of conserved cysteines in the protein, NLPs can be divided in types I, II, and III (Gijzen and Nürnberger, 2006; Oome and Van Den Ackerveken, 2014). Type I NLPs is the most frequent type; they have two conserved cysteines that are important for the

formation of a disulfide bridge (Gijzen and Nürnberger, 2006). Type II NLPs have four cysteines and formed two disulfide bridges (Gijzen and Nürnberger, 2006; Oome and Van Den Ackerveken, 2014). Type III NLPs usually have six cysteines residues and has been described in some ascomycetes and bacteria (Oome and Van Den Ackerveken, 2014; Seidl and Van den Ackerveken, 2019). The broad distribution of the three types in the Ascomycota have raised the question whether NLPs have originated within this phylum (Oome and Van Den Ackerveken, 2014). We previously investigated the evolutionary history of NLPs across the Dothideomycetes, the largest class of ascomycetes, and we found that the NLP superfamily has consistently small size among most of the species of this class of fungi (Chapter 2). Moreover, we found that the type I NLPs comprises a single family, was the most predominant across a variety of taxa, consisted of secreted proteins in most cases, and could be divided into three subfamilies: NLP1.1, NLP1.2, and NLP1.3 (Chapter 2). Amongst 33 species of Dothideomycetes, *Corynespora cassiicola* was the only species to carry a single NLP from each of the three type I NLP subfamilies (Chapter 2).

Corynespora cassiicola (Berk. & M. A. Curtis) C. T. Wei is usually referred as a necrotrophic pathogen due the type of lesions it causes on plants (Lopez et al., 2018). However, some isolates can be classified as endophytes and others as saprophytes (Déon et al., 2014, 2012b; Lopez et al., 2018). *Corynespora cassiicola* is one of the most destructive pathogens of the rubber tree (*Hevea brasiliensis*) in East Asia; causing the *Corynespora* Leaf Fall (Déon et al., 2012a; Lopez et al., 2018). On many host species, the disease is referred to as ‘target spot’ because the pathogen causes necrotic circular lesions with a cracked center that resemble to a target (Dixon et al., 2009; Sumabat et al., 2018). Target spot induces premature defoliation in a number of crops, which in turn reduces yield and results in economic losses. *Glycine max* (soybean), *Gossypium hirsutum* (cotton), *Cucumis sativus* (cucumber), and *Carica papaya* (papaya) are some of the plant hosts with economic importance to which *C. cassiicola* causes diseases (Déon et al., 2014; Dixon et al., 2009; Gao et al., 2020; Sumabat et al., 2018). In rare cases, *C. cassiicola* has been associated to human infection (Looi et al., 2017).

Corynespora cassiicola may comprise several phylogenetic lineages, each of which presenting different degrees of association with host specialization, pathogenicity and growth rate, but not with the geographical origin of the isolates (Déon et al., 2014; Dixon et al., 2009; Lopez et al., 2018). Levels of host specialization may

suggest the involvement of specialized effectors recognized by a limited range of compatible hosts (Gao et al., 2020; Lopez et al., 2018). So far, the necrotrophic toxin cassiicolin is the only completely characterized effector of *C. cassiicola* (Breton et al., 2000; de Lamotte et al., 2007; Déon et al., 2014; Lopez et al., 2018). Cassiicolin is an accessory effector, which means it is lineage specific and does not occur in all *C. cassiicola* isolates (Déon et al., 2014; Lopez et al., 2018; Wu et al., 2018). In addition to the cassiicolin-encoding genes, approximately 90 other effector genes are differentially expressed by *C. cassiicola* during rubber leaves infection; a NLP encoding-gene is amongst those effector genes (Lopez et al., 2018). The role that NLP effectors may played and whether all NLPs members of *C. cassiicola* are involved during infection remain elusive.

Herein, we aimed to investigate the evolutionary history of NLPs within a species framework. We accessed the genomic resource of 44 isolates of *C. cassiicola* available in public repositories to explore the genealogy of NLP effector genes at the species level. Additionally, we accessed the expression profiles of the NLP-encoding genes of *C. cassiicola* in the early phase of a pathogen-soybean interaction. We addressed the following four questions: i) Is the NLP family size conserved among isolates of *C. cassiicola*?; ii) To which extent did the evolutionary history of *C. cassiicola* shaped the genealogy of the NLP-encoding genes?; iii) What evolutionary forces shaped the evolution of NLP effector genes?; iv) What is the expression pattern of the NLP gene family during the early phase of host colonization by *C. cassiicola*? Unveiling the processes that allow the evolution and diversification of NLP-encoding genes in *C. cassiicola* may provide important information for understanding the evolution of effector genes and pathogenicity profiles of this ubiquitous plant pathogen.

MATERIAL AND METHODS

Data assembly

We assembled both protein and coding-DNA sequence (CDS) data from 44 isolates of *C. cassiicola*. From GenBank, we downloaded genome sequences from 42 fungal isolates and predicted their proteins and associated CDS. We also included proteins sequences and CDS from additional two isolates: the isolate CC_29 (Chapter

1) and the isolate CCP (Lopez et al., 2018). Thus, the set contained 44 isolates of *C. cassiicola* and, hereafter, was referred to as “the Corca set” (Table S2).

De novo genome assembly and gene prediction

Genomic data from two isolates of *C. cassiicola* (India_Hevea and TScotton1) were available as raw reads only; therefore, they required *de novo* genome assembly prior to subsequent analyses. We evaluated the quality of the raw reads using FastQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Prior to genome assembly, we ran Trimmomatic v0.36 (Bolger et al., 2014) using the settings “ILLUMINACLIP:TruSeq3.fa:2:30:10”, “LEADING:3”, “TRAILING:3”, and “MINLEN:50” to remove adaptors and low quality (Q < 30) reads. *De novo* assembly was carried out using SPAdes v3.9.0 (Bankevich et al., 2012) for Illumina reads, with the parameters set as “-k 21,33,55,77” for isolate India_Hevea and “-k 21,33,55” for isolate TScotton1. Additionally, we set the parameter “--careful” for both isolates. We accessed the quality of genome assemblies using QUAST v5.0.2 (Mikheenko et al., 2018), with default parameters for eukaryotes. Finally, we assessed the completeness of core fungal orthologs using BUSCO v4.0.5 (Simão et al., 2015) based on the dataset fungi_odb10 with 758 core orthologs.

For those genomes we had downloaded from GenBank, gene annotation was conducted using Augustus v3.2.2 (Stanke and Morgenstern, 2005). For training Augustus, we used the pre-existing gene structures available for the isolate CCP. Training was conducted according to manual instructions. Subsequently, we ran Augustus using isolate CCP as ‘reference species’ with parameters “--codingseq”, “--protein”, and “--cds”.

Protein annotation and assemble of NLP homologues

The predicted proteomes of the Corca set was annotated using PfamScan with Pfam v32.0 (El-Gebali et al., 2019) and InterProScan v5.30.69 (Jones et al., 2014) with the following eight parameters: SMART-7.1, SUPERFAMILY-1.75, ProDom-2006.1, CDD-3.16, TIGRFAM-15.0, Pfam v31.0, Coils-2.2.1, and Gene3D-4.2.0.

The presence of a signal peptide (according to SignalP v4.1; Petersen et al., 2011) and the absence of transmembrane domains (according to TMHMM v2.0; Krogh et al., 2001) defined the proteins that we predicted to be part of the secretome.

Subsequently, TargetP (Emanuelsson et al., 2007) predicted the subcellular localization of the predicted secretome.

Finally, HMMER v3.2.1 (http://hmmer.org) predicted the NLP homologues (E-value < 0.001) using profile hidden markov models (HMMs) for NPP1 domain (PF05630) from the predicted proteomes. A protein was declared to be a NLP when the NPP1 domain was annotated by at least two out three softwares: PfamScan, InterproScan and HMMER. Within the pool of NLPs, a given protein was considered to be a NLP effector when it passes the following three tests: (a) SignalP predicted it harbors a signal peptide, (b) TMHMM predicted it has no transmembrane domain, and (c) TargetP predicted it was part of a secretoy pathway. When a given NLP failed any of these three tests, it was regarded as a "non-effector NLP".

Orthogroups and phylogenomic analyses

OrthoFinder v2.3.3 (Emms and Kelly, 2015) established orthologous relationships among members of the Corca set (Table S1). OrthoFinder calculated length and phylogenetic distance-normalized scores from an all-versus-all alignment using DIAMOND v0.9.24 (Buchfink et al., 2014) and identified reciprocal best normalized hits. Normalized scores above default thresholds were assigned to the orthogroup graph and subjected to the Markov clustering analysis in order to assume ortholog groups (orthogroups).

We used a Maximum-likelihood phylogeny approach to uncovered the phylogenetic relationships among the isolates of *C. cassiicola*. To reconstruct the phylogeny, we prepared a dataset with the single copy ortholog (SCO) proteins identified by OrthoFinder. Proteins were aligned using the L-INS-I method as implemented in MAFFT v7.453 (Katoh and Standley, 2013). trimAL v1.4.rev22 (Capella-Gutiérrez et al., 2009) trimmed the alignments, with the parameters "-gt 0.7", "-cons 80", and "-w 5". After trimming and concatenation, we obtained dataset 1.

We subjected dataset 1 as input to IQ-TREE v1.6.11 (Nguyen et al., 2015) to estimate both the best evolutionary model and the phylogenomic tree. According to the Bayesian information criteria (BIC), IQ-TREE standard model selection indicated JTT+F+I+G4 as the best fit model for dataset 1. Subsequently, we carried out each Maximum-Likelihood phylogenomic tree using 1,000 ultrafast bootstrap replicates and ten independent runs. The trees were visualized in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Bayesian phylogenies

We assembled the additional dataset 2, which contained the predicted NLPs (CDS) that were present in the Corca set. Bayesian phylogenetic analyses based on dataset 2 allowed us to reconstruct the phylogenetic relationships among NLPs in *C. cassiicola*.

Alingments were obtained using the L-INS-I method, as implemented in MAFFT v7.453. During alignment of NLP CDS sequences of *C. cassiicola*, we found out that one of the NLP-encoding genes contained an insertion of either 105 bp (isolate CC_29) or 153 bp (isolates 777AA and RUD). These insertions were used as query in local BLASTn (E-value < 1e-4) to search for homologous sequences within the *C. cassiicola* genomes we had available and within the repeat library we have built previously for isolate CC_29 (Chapter 1).

To find the best fit model, dataset 2 was input into jModelTest v2.4 (Nylander, 2004). According to BIC, the best fit model was GTR+I. Subsequently, analysis was conducted on MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003), using two simultaneous runs, one cold chain and seven heated chains in each run. The number of generations was five million and the temperature was 0.25. Trees were sampled every 5000 generations and the first 250 trees were discarded as burn-in samples. Convergence was diagnosed with a) the standard deviation of split frequencies at the end of each run (below 1.5%; according to the output from MrBayes) and b) the Effective Sample Size (ESS) of each parameter (above 200 for all statistics) in Tracer v1.7.1 (Rambaut et al., 2018). For each analysis, a 50% majority-rule consensus tree of the two independent runs was obtained with posterior probabilities (PP) that were equal to bipartition frequencies. The phylogenetic tree was visualized in FigTree v1.4.4.

Network analysis and measures of nucleotide diversity

Gene genealogies of predicted NLP encoding-genes of *C. cassiicola* were inferred using the median-joining network method (Bandelt et al., 1999), as implemented in Network v4.613 (Fluxus Technology Ltd.) with default parameters. Preliminary analyses indicated that sequence alignments among introns were difficult to achieve; therefore, we choose to carry out the analyses using CDS data.

From dataset 2, we inferred the genealogical relationships among the full set of predicted NLP genes (CDS). Afterwards, we made partitions within dataset 2 according

to the sub-clades we had recovered during the phylogenetic analyses (see previous section). Each of the sub-clades underwent a new haplotype network analysis. Finally, we used DNAsp v6 (Rozas et al., 2017) to estimate the following five measures of molecular diversity in each of the sub-clades: H, number of haplotypes; H_d, haplotype diversity; π, nucleotide diversity; K, average number of nucleotide differences; S, number of polymorphic sites.

Selection analyses

We applied five different statistical tests to investigate how selection may have been shaping the NLP effector genes of *C. cassiicola*. The neutrality tests of Tajima's D (Tajima, 1989), Fu and Li's D* and F* (Fu and Li, 1993) were carried out in DNAsp v6. They tested for departures from neutrality based on allele frequency distribution indices (Ramírez-Soriano et al., 2008). As input for the analyses, we used each partition of dataset 2, according to sub-clades of NLP effector genes of *C. cassiicola.* For those analyses, we removed the information that would give rise to the signal peptides.

The McDonald and Kreitman test (MKT) (McDonald and Kreitman, 1991) tested the hypothesis of positive selection in the predicted NLP effector genes of *C. cassiicola*. The MKT test calculates the ratio of the number of non-synonymous polymorphic sites (*Pn*) by the number of synonyms polymorphic sites (*Ps*) within the species compared to the ratio of the number of non-synonymous nucleotide substitutions (*Dn*) by the number of synonymous nucleotide substitutions (*Ds*) between species; thus, an outgroup is required to determine in which sites the differences are fixed (Rody and Oliveira, 2018). The MKT calculated the Neutrality Index (NI), which indicates how far polymorphism is from neutral evolution. Under neutral evolution, NI values are expected to be equal to 1. If NI < 1, there is an excess of fixation of non-neutral substitutions and it indicates positive selection; if NI > 1 suggests that negative selection is removing harmful mutations. Additionally, the MKT calculated the rate of positive selection (α). The parameter can vary from -∞ to 1. Positive values of α corroborates with positive selection. The MKT analyses were carried out in DNAsp, using the partitions we had obtained for dataset 2. The MKT analyses required outgroups. Therefore, we returned to the results of the Bayesian phylogenetic analyses of NLPs from Dothideomycetes we performed previously (Chapter 2) and selected as

outgroup the closest orthologue protein to each of the NLPs of *C. cassiicola*. For those analyses, we removed the information that would give rise to the signal peptides.

The package *codeml* from PAML v4.9j (Yang, 2007) carried out tests of positively selected sites in the NLP effectors of *C. cassiicola*. The package calculates the site-to-site variation, the ratio of nonsynonymous (d*N*, amino acid changing) to synonymous (d*S*, amino acid retaining) substitution rates (ω = d*N*/d*S*). Sites that are subject to natural selection show dN/ dS = 1, whereas those that are under purifying selection and those that are evolving by positive selection have dN/dS < 1 and dN/dS > 1, respectively (Terauchi and Yoshida, 2010). To ensure that ω variation represented amino acid sites that were fixed along independent lineages, we used the same inputs as for MKT. Possible ambiguities and alignment gaps were cleanup by choosing option cleandata = 1. A phylogenetic tree for each NLP effector gene was generated by PhyML v3.1 (Guindon and Gascuel, 2003), using the GTR model of nucleotide substitution, and used as input along with the aligned datasets. We performed likelihood ratio tests (LTRs) between two pairs of models: M2 (selection) against M1 (neutral) and M8 (beta&ω) against M7 (beta). The positively selected codon sites were predicted using Bayes Empirical Bayes (BEB) method. Condon sites were considered under positive selection when ω > 1 and posterior probability calculated by the BEB was > 95%.

Motif searching in the NLP effectors of *C. cassiicola*

The packages GLAM2SCAM and GLAM2, as implemented in the MEME suite v5.1.0 (Bailey et al., 2009), created a consensus logo of the NLP effectors in *C. cassiicola*. The package GLAM2 searched for pattern of motifs among predicted NLP effectors; while GLAM2SCAM searched for matches for the motif initially discovered by GLAM2. Each match received a score, indicating how well it fit the motif. GLAM2 returned a logo with the consensus sequence with the highest score.

Visualization of the 3D structure of the NLP effector was obtained in PyMOL v2.3 (Schrödinger, LLC), using as input the PBD format file from *Pythium aphanidermatum* that was available under accession 3GNU at the Protein Data Bank. The protein sequence was depicted as a cartoon diagram and color coded. In the cartoon, the sites that *codeml*/PAML suggested to be under positive selection were highlighted.

Expression of the NLP genes of *C. cassiicola*

Using RT-qPCR, we investigated the expression *in planta* of the predicted NLP genes of *C. cassiicola*. We used *C. cassiicola* isolate CC_29 (originally isolated from soybean leaves) because it also contained a copy of the rare non-effector Cc_NLP4 and its genome had been characterized previously (Chapter 1).

Leaves of soybean plants cultivar "MG/BR-46 Conquista" were inoculated when they reached stage V3. Inoculations were as follows: four 20 µl droplets of conidia suspension (3.5 x 10^4 conidia/ml) from isolate CC_29 amended with 0.01% Tween 20 were spotted on the abaxial face of each leaflet from trifoliate leaves. Plants were incubated in a humid chamber at 25 ± 2 °C. Controls consisted of leaves inoculated with sterile distilled water supplemented with 0.01% Tween 20. Sampling consisted of the collection of four discs (1.4 cm^2) per leaflet (at the inoculation spots) using a cork borer at 0, 6, 12, and 18 hours post inoculation (hpi). For each time point, three independent biological repeats were collected, each of which from a distinct plant. Samples were immediately frozen in liquid nitrogen and stored at -80°C until RNA extraction.

Total RNA was extracted from each sample using TRIzol reagent (Life Technologies) according to manufacturer's specifications. RNA was quantified on a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific) and visualized on 1.5% (w/v) agarose gel. A total of 4 µg of RNA was treated with one unit of DNase I Amplification Grade (Invitrogen) to remove potential contamination with genomic DNA. The cDNA synthesis was performed using 4 µg of total RNA, Oligo$(dT)_{18}$ primers, and the M-MLV Reverse Transcriptase (Invitrogen) following the manufacturer's instructions.

Using Primer Blast (https://www.ncbi.nlm.nih.gov/tools/primerblast/), we designed primers for each predicted NLP gene (CDS) and for the β-tubulin gene (CDS) from *C. cassiicola* isolate CC_29. We checked primer quality using NetPrimer (http://www.premierbiosoft.com/netprimer/). Primers used in this study are listed (supplementary Table S2).

Amplifications (RT-qPCR) were performed on a 7500 Real-Time PCR System (Applied Biosystems) using SYBR Green PCR Master Mix (Life Technologies), specific primers (Table S2), and 1 µl of a 2-fold diluted cDNA. The reactions were performed as follows: 2 min at 50 °C, 10 min at 95 °C, and 40 cycles of 94 °C for 15 seconds, and 60 °C for 1 min. To confirm quality and primer specificity, we visually inspected the T_m

(melting temperature) of amplification products in dissociation curves for each primer individually.

For quantitation of the NLP genes, we used β-tubulin as the endogenous control gene for data normalization. Gene expression was quantified using the $2^{-\Delta Ct}$ method for absolute quantification. To confirm the observed differences in gene expression, we applied a one-way ANOVA followed by Tukey's test (*p*-value < 0.05) for each NLP gene, independently. The analyses were implemented in R v3.6.2 (http://www.R-project.org/). A heatmap was designed using the *tidyr* and *ggplot2* packages, also in R.

RESULTS

Intraspecific genome-wide phylogeny in *C. cassiicola*

De novo genome assembly of isolates India_Hevea and TScotton1, available on GenBank as raw reads, resulted in good quality genomes (Table S3). The final genome assemblies of India_Hevea and TScotton1 added up to 42.89Mbp (arranged in 1,059 scaffolds; ≥ 500 bp) and 42.11Mbp (2,188 scaffolds; ≥ 500bp), respectively. The BUSCO analysis identified 99.6% of the orthologs as complete for India_Hevea, and 98.9% as complete for TScotton1. The BUSCO scores indicated a high level of completeness, making the genomes suitable for the downstream analyses.

Genomic data within the Corca set allowed OrthoFinder to uncover 9,018 SCO proteins (a concatenated set of 4,840,641 amino acids). The whole-genome data provided suppport for the phylogenetic reconstruction of isolates of *C. cassiicola* (Figure 1). The phylogenomic tree showed well supported nodes (bootstrap values = 100, for the main nodes). To a certain extent, clade composition lacked association with either host or geographic origin. The most member-rich clade contained 24 isolates from five host species (*Cucumis sativus, Glycine max, Gossypium hirsutum, Hevea brasiliensis,* and *Vernonia cinerea)* that were sampled in the Americas, Africa, and Asia. Second clade comprised 13 isolates from three host species (11 from *H. brasiliensis*, one from *Solanum lycopersicum*, and one from *Piper hispidinervum*) that were collected in the Americas, Africa, and Asia. The third clade contained seven isolates from two host species (six from *H. brasiliensis,* one from *Homo sapiens*) that were collected in Asia.

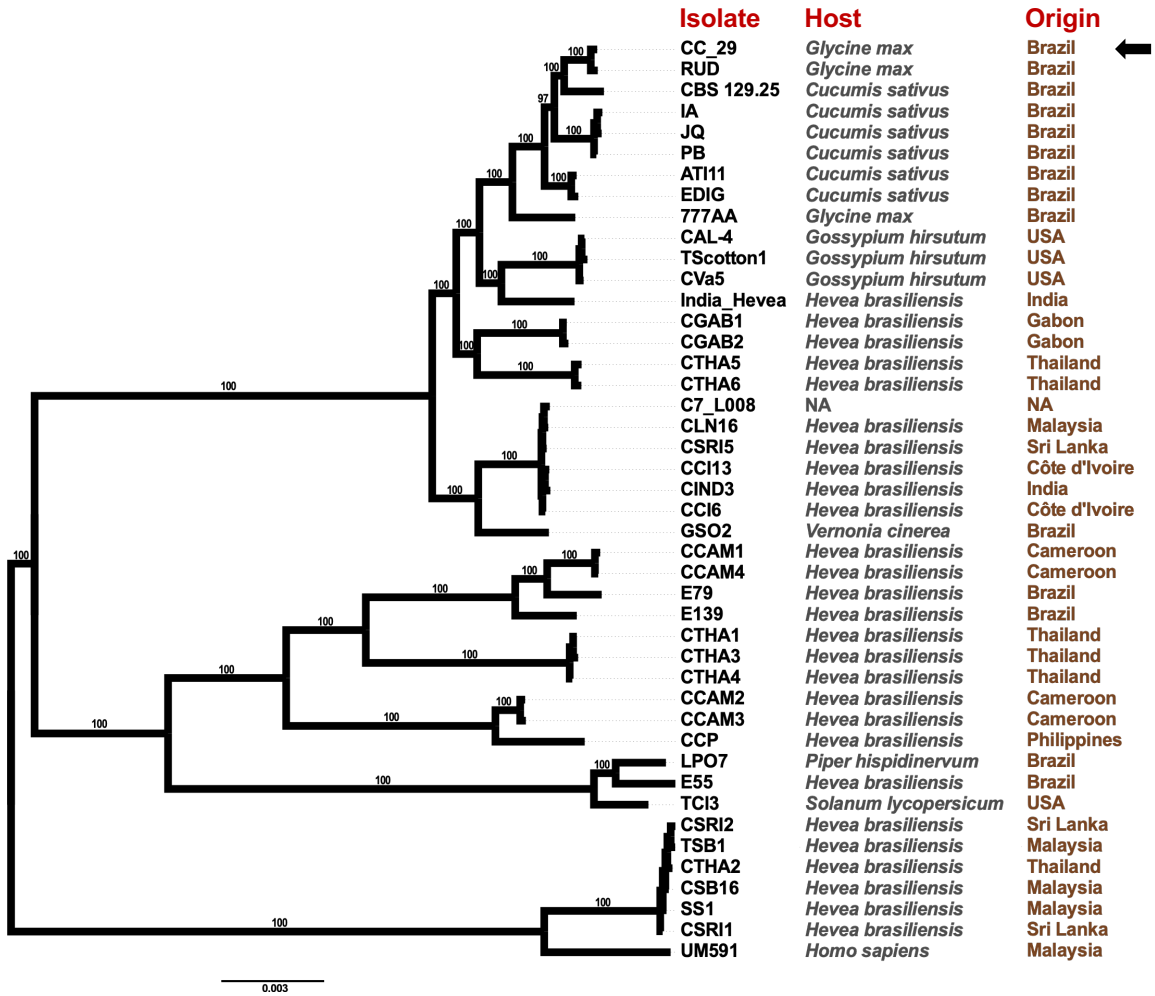


Figure 1. Phylogenomic relationships among 44 isolates of *Corynespora cassiicola*. The unrooted maximum likelihood consensus tree was based on a dataset of 9,018 single copy ortholog proteins. An arrow indicates the placement of isolate CC_29. Host species and geographical origin of each isolate are as indicated. Branch lengths are drawn to scale; nodal support values are given as local bootstraps above the branches. Scale bar corresponds to the expected number of substitutions per site.

Phylogenetic relationships among NLP-encoding genes of *C. cassiicola*

The genome sequences of 44 isolates of *C. cassiicola* yielded a total of 135 predicted NLP-encoding genes (CDS). The phylogenetic tree (Figure 2) showed well-supported nodes (posterior probabilities = 100, for all main nodes). The 135 members were grouped into four major clades. Based on OrthoFinder analysis, each of the four major clades corresponded to an orthogroup.

Within each of the three major clades, the members shared homology to one of the three NLPs we had previously uncovered in the genome of the isolate CCP (Corca_601720, Corca_537044, and Corca_497672; Chapter 2). Therefore, we named those three major clades after the three sub-clades within the type I NLP clade we had recovered previously (Chapter 2): Cc_NLP1.1, Cc_NLP1.2, and Cc_NLP1.3 (Figure 2). Every isolate had one copy of NLP-encoding gene within each of the three clades. Those 132 genes were predicted to be NLP effector-encoding genes.

In addition to the three widespread copies, however, we predicted a fourth copy in the genomes of three isolates obtained from soybean that was cultivated in Brazil: CC_29 (gene CC_29_g14222), 777AA (777AA_g14294), and RUD (RUD_g11851). In the phylogenetic tree (Figure 2), the additional copies clustered together to formed the fourth major clade (hereafter referred to as Cc_NLP1.4), which exhibited a sister relationship to Cc_NLP1.3.

Members of the Cc_NLP1.4 clade showed features that distinguish them apart from the members of the other three major clades. In addition to harboring many exclusive synonymous and non-synonymous substitutions, they carry a large insertion at the 5’ region, just after the predicted start codon (105 bp in CC_29_g14222; 153 bp in both 777AA_g14294 and RUD_g11851). BLASTn searches using each of those insertion sequences as queries did not find significant matches in the genomes of the remaining 41 isolates of *C. cassiicola* (Corca set). Moreover, additional BLASTn searches did not find evidence for homology between those insertions and the repeat/transposable elements that are present in the genome of isolate CC_29. Finally, members of the Cc_NLP1.4 clade were predicted to be non-effector NLP-encoding genes.

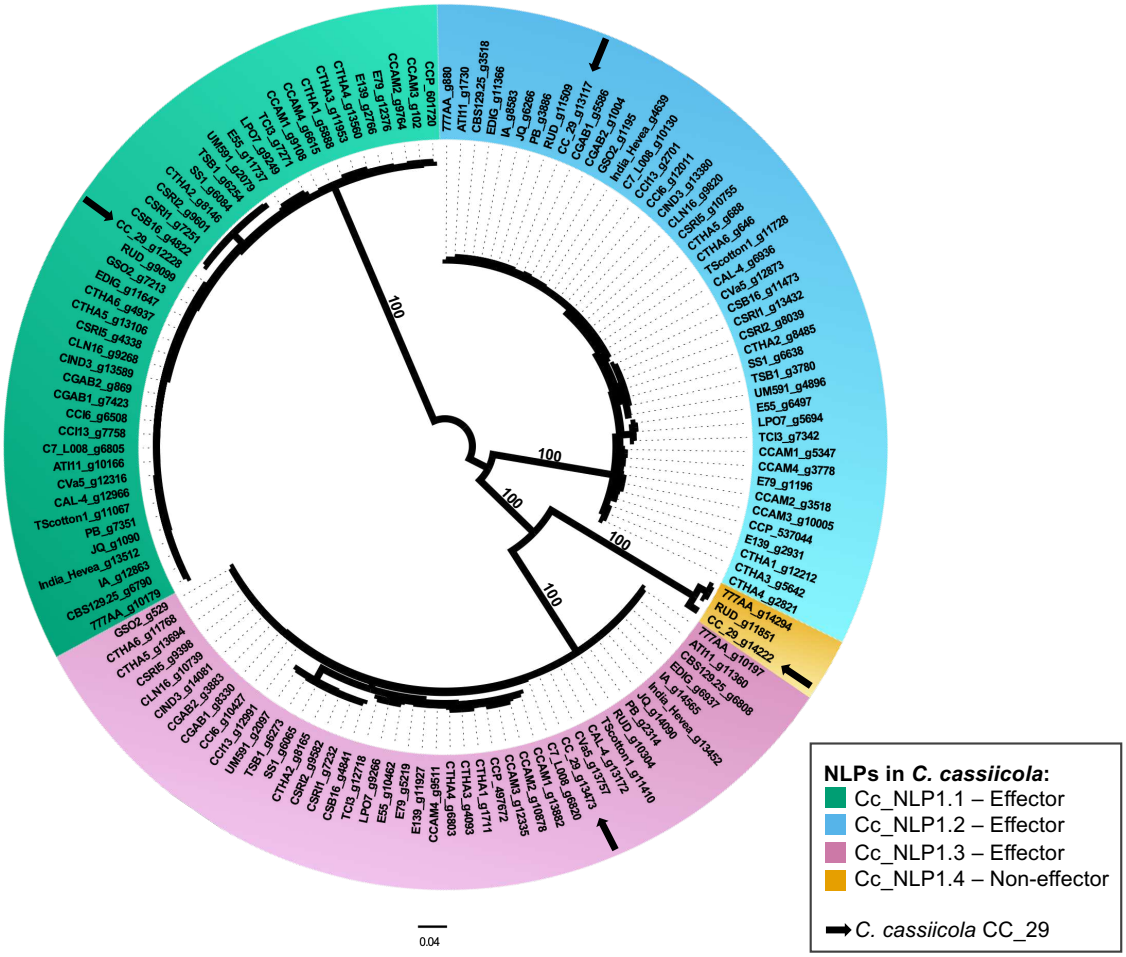


Figure 2. Phylogenetic relationships among 135 predicted Necrosis- and Ethylene-inducing peptide 1-like protein (NLP)-encoding genes (CDS) in *Corynespora cassiicola*. The DNA dataset for the unrooted Bayesian phylogeny (consensus tree) was 900 base pair long and contained NLP of 44 isolates. Arrows indicate the placements of the four predicted NLP-encoding genes of *C. cassiicola* isolate CC_29. Along the tree, gene ID according to gene prediction using Augustus (or MycoCosm nomenclature for isolate CCP) followed the name of the isolate and identified the predicted NLP-encoding genes. Gene names are color-coded according to the CDS homology to any of the three effector NLPs or to one non-effector NLP (Cc_NLP1.4) of *C. cassiicola* CC_29, as indicated. Branch lengths are drawn to scale; nodal support values are given as posterior probabilities (when ≥ 95%) above the branches. Scale bar corresponds to the expected number of substitutions per site

Gene genealogies among NLP-encoding genes of *C. cassiicola*

The full haplotype network recovered 39 haplotypes among the 135 sequences of NLP-encoding genes (CDS) that had been retrieved from genomes of 44 isolates of *C. cassiicola* (Figure 3). The network exhibited four major sets of haplotypes (hereafter, each major set is referred to as a 'haplogroup'). Without exception, both the topology of the network and the composition of each haplogroup were congruent with the four clades we had obtained in the preceding phylogenetic analysis (Figure 2). We named the four haplogroups after the four phylogenetic clades: Cc_NLP1.1, Cc_NLP1.2, Cc_NLP1.3, and Cc_NLP1.4. While Cc_NLP1.1 and Cc_NLP1.4 occupied each a tip position on opposed ends of the full network, Cc_NLP1.2 and Cc_NLP1.3 occupied its center. The largest number of mutational steps between nearest haplogroups occurred between Cc_NLP1.1 and Cc_NLP1.2 (314 steps); while the smallest number occurred between Cc_NLP1.3 and Cc_NLP1.4 (234 steps). Distances between nearest haplogroups far exceeded the distance between nearest haplotypes within haplogroups. Therefore, we split the full dataset according to haplogroups and analyzed each haplogroup independently.

The four partial networks shed further light on the genealogical relationships among haplotypes within each of the haplogroups. Within each haplogroup, the mutational distances that separated two nearest haplotypes were short (up to 11, within Cc_NLP1.3; up to 24 within Cc_NLP1.4). Mutational distances larger than two arise owing to missing intermediates. Those missing intermediates represent either extinct haplotypes or, more likely, extant haplotypes that may exist in the population of isolates yet to be sampled. Within each haplogroup, most of the missing intermediates occurred in tandem. The presence of large number of autapomorphic substitutions suggested that most of the NLP-encoding genes have experienced relatively high levels of genetic differentiation since their origin.

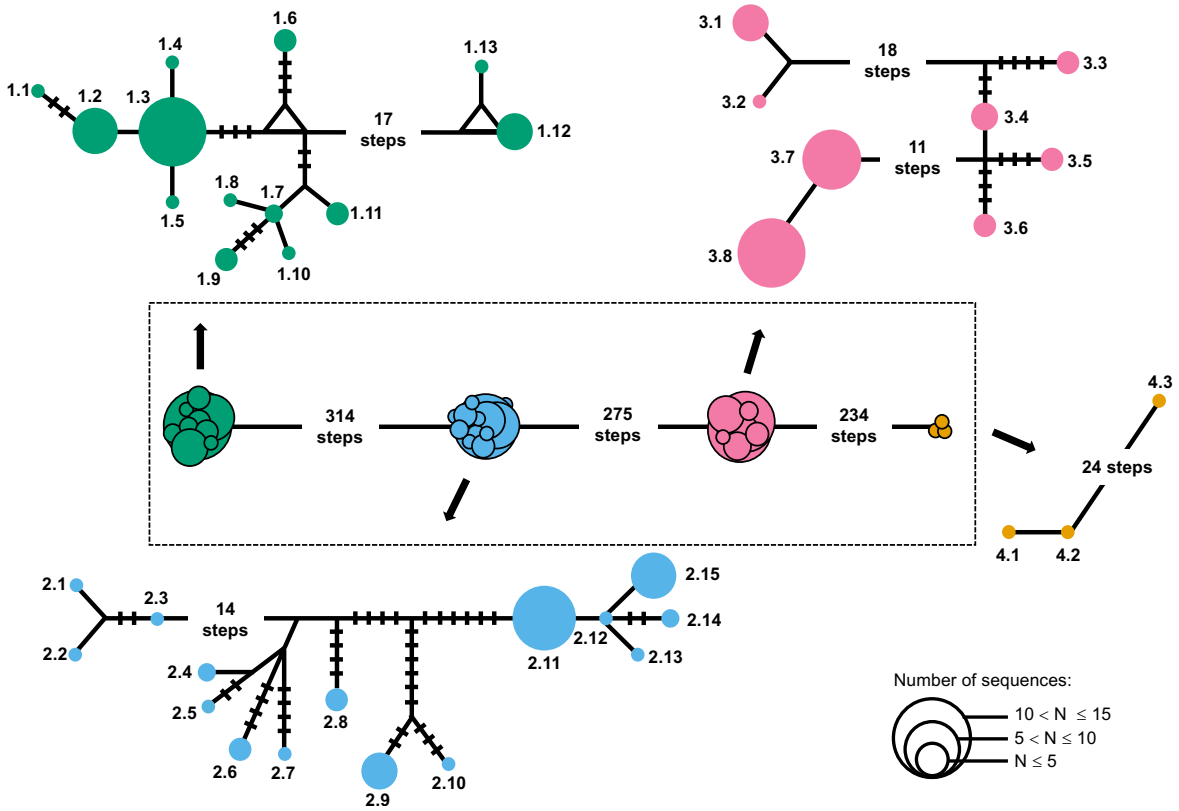


Figure 3. Genealogical relationships of the four haplogroups of predicted Necrosis- and Ethylene-inducing peptide 1-like protein (NLP)-encoding genes (CDS) of *Corynespora cassiicola*. Median-joining networks of each of the four haplogroups are color-coded according to the predicted NLP encoding-genes: green, Cc_NLP1.1; blue, Cc_NLP1.2; pink, Cc_NLP1.3; and orange, Cc_NLP1.4. A central box depicts the full set of haplogroups; the number of mutational steps is as indicated. A circle represents a given haplotype (coded with numbers); circle size is proportional to the relative frequencies as indicated. Numbers of mutational steps are indicated with bars when more than one (unless indicated otherwise).

Measures of nucleotide diversity indicated that the four haplogroups of NLP-encoding genes of *C. cassiicola* followed relatively distinct evolutionary paths (Table 1). Genetic differentiation reached the lowest values among member of haplogroup Cc_NLP1.1, with only 9 mutations as the average number of nucleotide differences (K) among pair of haplotypes. Thus, haplotype (H_d) and nucleotide diversity (π) were as also lower among members of Cc_NLP1.1. On the opposite end, genetic differentiation reached the highest values among members of haplogroup Cc_NLP1.4, in which pairs of haplotypes diverged by about 17 mutations on average, showing substantially large values of haplotype (H_d) and nucleotide diversity (π).

Table 1. Measures of molecular diversity for the four haplogroups of NLP encoding-genes (CDS) in *Corynespora cassiicola*.

NLP Haplogroup	# Sequences	# Variable Sites (S)	# Haplotypes (H)	Haplotype Diversity (H_d)	Nucleotide Diversity (π)	Average # of Nucleotide Differences (K)
Cc_NLP1.1	44	35	13	0.86	$1.3x10^{-2}$	8.97
Cc_NLP1.2	44	57	15	0.88	$1.6x10^{-2}$	11.68
Cc_NLP1.3	44	39	8	0.83	$1.6x10^{-2}$	12.01
Cc_NLP1.4	3	25	3	1.00	$2.2x10^{-2}$	16.67

Selection analyses on NLP effector genes in *C. cassiicola*

We applied three selective neutrality tests (Tajima's D, Fu and Li's D* and F*) taking into consideration three sets of NLP effector-encoding genes (Cc_NLP1.1, Cc_NLP1.2, and Cc_NLP1.3) (CDS) of *C. cassiicola* (Table 2, Figure S1). Independent selective tests were carried out using both the full extension of the genes and a contiguous sliding window of 25 bases. The Tajima's D test recovered non-significant values for each of the three sets (*p*-value > 0.10), for both the full sequence extension and for the contiguous sliding window. Thus, results of the Tajima's D test supported the null hypothesis of neutrally evolving genes.

The Fu and Li's D* and F* tests provided similar results. The F* statistics are shown in Figure S1. Using the full extension of the genes, Fu and Li's D* and F* recovered non-significant values for two sets (Cc_NLP1.1 and Cc_NLP1.2), with *p*-values > 0.10; and significant positive values for Cc_NLP1.3 (*p*-value < 0.02 for D*; *p*-value < 0.05 for F*) (Table 2). Therefore, the D* and F* tests rejected the null hypothesis of neutrally evolving genes and suggests that balancing selection is acting

upon Cc_NLP1.3. Using the sliding window approach, Fu and Li's D* and F* recovered significant negative values for some nucleotides between the midpoints 616-650 of Cc_NLP1.2 (*p*-value < 0.10 for D*; *p*-value < 0.05 for F*) (Figure S1).

To test for positive selection, we applied the independent MKT on the full extension of each of the three sets of NLP effector-encoding genes (Cc_NLP1.1, Cc_NLP1.2, and Cc_NLP1.3) (CDS) of *C. cassiicola.* The MKT requires outgroup; therefore, we added the following homologous sequences (chosen based on the results of our previous phylogenetic study on NLPs across the Dothideomycetes (Chapter 2): Cc_NLP1.1 (Stano_3364 from *Stagonospora nodorum*), Cc_NLP1.2 (Bimnz_484389 from *Bimuria nova-zelandiae*), and Cc_NLP1.3 (Cocsa_135644 from *Cochliobolus sativus*). Results of the MKTs showed significant values for Cc_NLP1.1 (*p*-value < 0.01) and Cc_NLP1.2 (*p*-value < 0.001), according to Fisher's exact test and G-test, as implemented in DNAsp. However, results showed NI < 1 and α values $0 < \alpha < 1$ (Table 2) for all NLP effector genes (Cc_NLP1.1, Cc_NLP1.2, and Cc_NLP1.3). Together, these results indicate an excess of fixation of non-neutral substitutions on NLP effector genes and suggest that, to some extent, they evolved under positive selection.

To test for selective pressure upon amino acid sites, we evaluate positive selection using site-to-site models (M1 vs. M2a, M7 vs. M8), as implemented in the *codeml* tool from PAML software package. Congruent results were obtained for models M2 and M8 (both for selection). The results of the M8 model are shown (Table 2). In the Cc_NLP1.3, two sites were identified to be under positive selection (PP > 95%) as calculated by the BEB approach (Figure 4). Positively selected sites referred to residue 36, which was an arginine (R36), and residue 77, which was an asparagine (N77). Both residues are located on regions that are predicted to be α-helices, located at one side of the protein and adjacent to the negatively charged (cation binding) cavity that forms around the conserved GHRHDWE motif. The alternative allele for the positive selected residue 36 detained a lysine (K36), instead of an arginine (R36). For the positive selected residue 77, alternative alleles detained either a glycine (G77) or a serine (S77), instead of an asparagine (N77).

Table 2. Results of neutrality tests for three NLP effector genes (CDS) in *Corynespora cassiicola*.

Effector Gene	Tajima's	Fu and Li's		McDonald and Kreitman			Positive selected amino acid sites (ω > 1) (amino acid)[2]
	D (*p*-value)	D*[1] (*p*-value)	F*[1] (*p*-value)	NI	α	*p*-value	
Cc_NLP1.1	0.01 ($p > 0.10$)	0.79 ($p > 0.10$)	0.62 ($p > 0.10$)	0.16	0.84	$p < 0.01$	-
Cc_NLP1.2	-0.47 ($p > 0.10$)	0.26 ($p > 0.10$)	0.01 ($p > 0.10$)	0.17	0.83	$p < 0.001$	-
Cc_NLP1.3	1.15 ($p > 0.10$)	1.63 ($p < 0.02$)	1.73 ($p < 0.05$)	0.52	0.48	$p > 0.10$	36 (R)*, 77 (N)**

[1] Asterisks indicate no outgroup.
[2] The *codeml* tool performed predictions with model M8, as implemented in the PAML 4.9j software package. BEB calculated the Bayesian PPs (>95%* or >99%**).

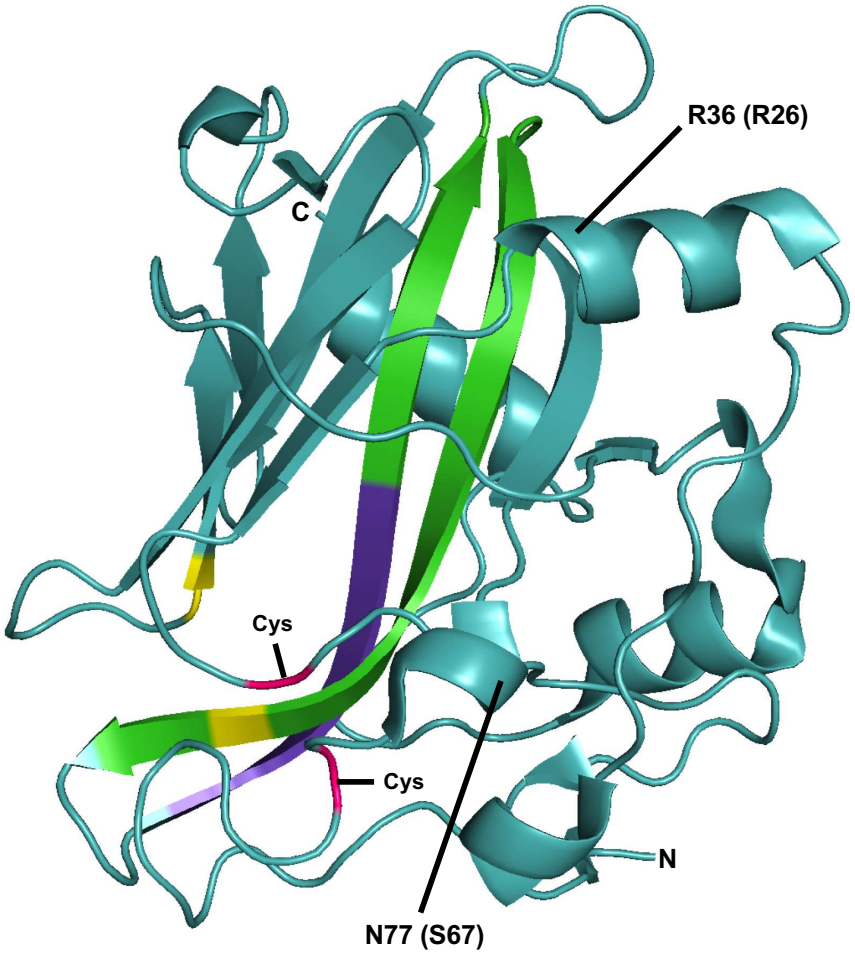


Figure 4. PyMOL diagrams depicting the 3D structure of the Necrosis- and Ethylene-inducing peptide 1-like (NLP) effector of *Pythium aphanidermatum* and the relative locations of two positively selected sites of the Cc_NLP3 effector in *Corynespora cassiicola*. Protein structure was taken from the Protein Data Bank (accession number 3GNU). According to *codeml*/PAML, there were two positively selected amino acid sites (R36 and N77) in Cc_NLP1.3. R26 and N67 are the corresponding sites in the structure of the NLP effector of *Pythium aphanidermatum*. Color codes: purple, conserved GHRHDWE site; green, semiconserved sites in central β-strands; yellow, residues directly involved in cation binding; pink; conserved cysteine residues involved in the protein stabilization (disulfide bridge formation). Letter codes: Cys, conserved cysteines residues; N, N-terminus; and C, C-terminus.

Nature and context of nucleotide substitutions within NLP effector-encoding genes in *C. cassiicola*

We documented the nature and context of nucleotide substitutions along the haplotypes within haplogroups Cc_NLP1.1, Cc_NLP1.2, and Cc_NLP1.3 (Figure S2). During each independent analysis, the respective gene member we had identified in isolate CC_29 was taken as the reference sequence: CC_29_g12228 (Cc_NLP1.1), CC_29_g13117 (Cc_NLP1.2), and CC_29_g13473 (Cc_NLP1.3).

Haplogroup Cc_NLP1.1 contained a total of 35 nucleotide substitutions among its 13 haplotypes (Figure S2a, Table 1). Among them, four (10%) were non-synonymous mutations, that is, they resulted in amino acid replacements. One of them was a conservative amino acid replacement at the second half of the NPP1 domain of haplotypes 11-13. The remaining three mutations (corresponding to one conservative and two non-conservative amino acid replacements) were located outside the NPP1 domain.

Haplogroup Cc_NLP1.2 exhibited the highest number of nucleotide substitutions; there were 57 of them among its 15 haplotypes (Figure S2b, Table 1). Among them, eight (14%) were non-synonymous and resulted in seven distinct amino acid replacements in the protein sequence. Two consecutive non-synonymous mutations resulted in a non-conservative amino acid replacement at the first half of the NPP1 domain, adjacent to the GHRHDWE motif. The remaining non-synonymous mutations corresponded to conservative amino acid replacements: five of them in the second half of the NPP1 domain and one immediately after the domain.

Among the eight haplotypes of haplogroup Cc_NLP1.3, there was a total of 39 nucleotide substitutions (Figure S2c, Table 1). Amongst the substitutions, eight (20%) were non-synonymous and resulted in seven different amino acid replacements in the protein sequence. At the first half of NPP1 domain, a non-synonymous mutation resulted in conservative amino acid replacement (haplotypes 1-4). Also, in the first half of the NPP1 domain, two consecutive non-synonymous mutations resulted in three different amino acids replacements at the same codon (haplotypes 1-3, 5). According to the *codeml*/PAML analysis, the sites that harbor these amino acid replacements (conservative and non-conservative) corresponded to positively selected sites (Figure 4, Table 2). The remaining non-synonymous mutations (five) corresponded to conservative amino acid replacements: one took part in the CDS portion encoding for the signal peptide, followed by one in the portion encoding between the signal peptide

and the NPP1 domain. Finally, three conservative amino acid replacements took part in the CDS portion encoding for the NPP1 domain (one in the first half and two in the second half of the domain).

Expression patterns of NLP-encoding genes of *C. cassiicola* isolate CC_29

Using soybean as a host species and RT-qPCR analyses, we characterized the expression patterns of the four NLP-encoding genes of *C. cassiicola* isolate CC_29. According to our previous analyses, the genes CC_29_g12228 (Cc_NLP1.1), CC_29_g13117 (Cc_NLP1.2), and CC_29_g13473 (Cc_NLP1.3) were predicted to be NLP effector genes, while CC_29_g14222 (Cc_NLP1.4) was predicted to be a non-effector NLP.

During the first 18 hours post inoculation, the NLP effector genes showed varying levels of expression (Figures 5 and S3). The effector gene CC_29_g12228 (Cc_NLP1.1) showed the earliest and the highest level of absolute expression (fold variation). The expression of CC_29_g12228 gene (Cc_NLP1.1) experienced a significant increase between 12 and 18 hpi. Although expression of the CC_29_g13117 gene (Cc_NLP1.2) remained always very low during the time-points evaluated, there was a significant increase at 18 hpi. The effector gene CC_29_g13473 (Cc_NLP1.3) had lower levels of expression compared to the CC_29_g12228 gene (Cc_NLP1.1) and higher levels of expression compared to CC_29_g13117 gene (Cc_NLP1.2), with levels of expression increasing from 12 hpi onwards.

The non-effector NLP gene (CC_29_g14222; Cc_NLP1.4) showed low (almost undetectable) levels of expression during the time-points evaluated. However, its expression was significantly higher at 0 hpi.

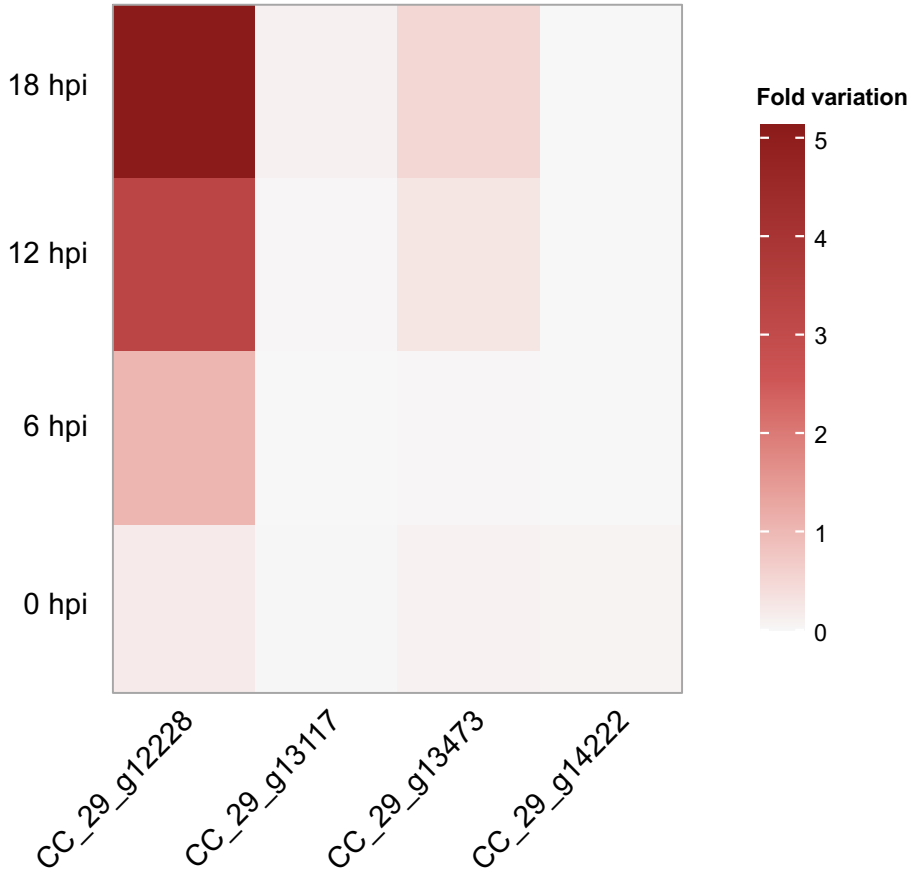


Figure 5. Expression patterns of the Necrosis- and Ethylene-inducing peptide 1-like (NLP) genes of *Corynespora cassiicola* (isolate CC_29) after inoculation in soybean plants. Gene expression was inferred by RT-qPCR analyses. The fold variation of gene expression of each NLP gene was calculated using the $2^{-\Delta Ct}$ method, the values were normalized to an endogenous control gene (β-tubulin). Columns correspond to NLP genes (from left to right): Cc_NLP1.1 (CC_29_g12228), Cc_NLP1.2 (CC_29_g13117), Cc_NLP1.3 (CC_29_g13473), and Cc_NLP1.4 (CC_29_g14222). Rows correspond to time in hours post inoculation (hpi) (from bottom to top): 0, 6, 12, and 18. Scale bar corresponds to fold variation. Heatmap generated based on Figure S3.

Overview on the NLP effector proteins of *C. cassiicola*

Based on our annotation analyses, we proposed a scheme to illustrate the protein structure of an NLP effector in *C. cassiicola* (Figure 6a). The sizes of the effector proteins varied from 237 (Cc_NLP1.1) to 243 (Cc_NLP1.3) amino acids long. The protein consisted of a signal peptide (predicted by SignalP) and a single NPP1 domain (predicted by HMMER). The sizes of the signal peptides varied from 18 (in Cc_NLP1.2 and Cc_NLP1.3) to 21 (in Cc_NLP1.1) amino acids long. Sizes of the NPP1 domains varied from 185 (in Cc_NLP1.1) to 190 (in Cc_NLP1.2 and Cc_NLP1.3) amino acids long. A segment varying from 27 (in Cc_NLP1.1) to 33 (in Cc_NLP1.3) amino acids connected the signal peptide to the NPP1 domain. After the NPP1 domain, there were additional two (in Cc_NLP1.2 and Cc_NLP1.3) to four (in Cc_NLP1.1) amino acids at the carboxy-terminus of the protein. The two conserved cysteine residues that play crucial role in protein stabilization are located at the amino-terminal side of the NPP1 domain. The positively selected sites that *codeml*/PAML analysis had identified are also located in the initial part of NPP1 domain of the NLP effectors of the Cc_NLP1.3 haplogroup.

A consensus pattern for the three NLP effectors (Cc_NLP1.1, Cc_NLP1.2, and Cc_NLP1.3) revealed the conserved GHRHDWE motif in the central region of the protein (positions 29-35, Figure 6b). This conserved motif is surrounded by two semiconservative sites (positions 11-24 and 36-40). Together, this semiconserved region are part of two β-strands in the central portion of the protein (Figure 4). One residue involved in the cation binding (aspartic acid, D) is conserved among the three NLP effectors (position 22, Figure 6b) in *C. cassiicola* and located in the negatively charged (cation binding) cavity of the protein secondary structure (Figure 4).

The non-effector Cc_NLP1.4 diverged greatly in amino acid sequence from the NLP effectors (Cc_NLP1.1, Cc_NLP1.2, and Cc_NLP1.3). The highest level of sequence similarity was found between Cc_NLP1.4 and Cc_NLP1.3 (49%) and the lowest between Cc_NLP1.4 and Cc_NLP1.1 (30%). This high level of similarity between Cc_NLP1.4 and Cc_NLP1.3 is congruent with the placement of Cc_NLP1.4 in the phylogenetic analysis (Figure 2) and network analysis (Figure 3). In Cc_NLP1.4, there was a tyrosine (Y) at the fourth residue of the heptapeptide motif (GHRYDWE) instead of the histidine (H) found in the motif of NLP effectors (GHRHDWE, Figure 6b). Moreover, the NPP1 domain of Cc_NLP1.4 was shorter compared to those predicted for the NLP effectors; it consisted of 109 amino acids and started immediately from the

GHRYDWE motif. Thus, semiconserved sites that anticipate the heptapeptide motif in NLP effectors were not present in the non-effector Cc_NLP1.4 of *C. cassiicola*. Other residues of Cc_NLP1.4 included 142 (CC_29_g14222) to 159 (777AA_g14294, and RUD_g11851) amino acids upstream the NPP1 domain and one amino acid downstream the domain.

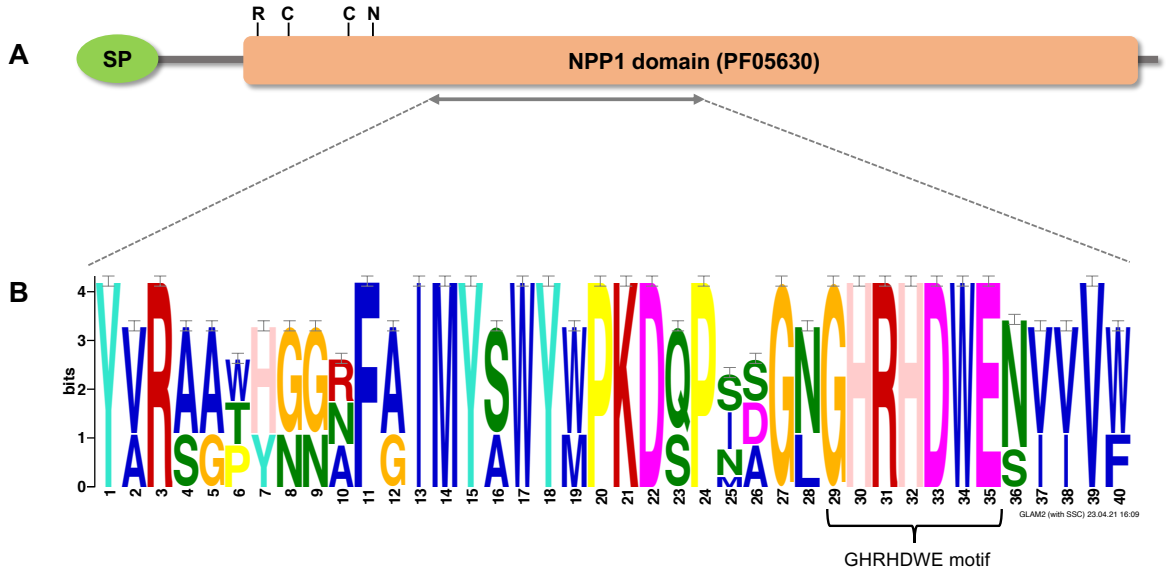


Figure 6. Overview of the Necrosis- and Ethylene-inducing peptide 1-like (NLP) effector proteins of *Corynespora cassiicola*. A. SignalP predicted the signal peptide (SP); HMMER predicted the NPP1 domain (PF05630). Horizontal grey segments depict the remaining amino acid residues. Over the NPP1 domain: C, conserved cysteine residues; R36 and N77, two positively selected residues of the Cc_NLP3 effector (see Figure 4). **B.** Consensus pattern within the three NLP effectors (Cc_NLP1.1, Cc_NLP1.2, and Cc_NLP1.3). The variation pattern of amino acids surrounding the conserved GHRHDWE motif (positions 29-35) was generated with GLAM2SCAN and was visualized with GLAM2. Positions 1-3, 11-24, and 36-40 depict semiconserved sites. Position 22 contains a conserved residue of aspartic acid that is directly involved in cation binding.

DISCUSSION

Phylogenetic relationships among isolates of *C. cassiicola*

Genetically closely-related isolates were obtained from widely distant locations worldwide, suggesting *C. cassiicola* took advantage of long-distance dispersal events to reach new territories (Dixon et al., 2009). The wide range of host species and agriculturally-mediated events are likely driving forces that allow for the spread of *C. cassiicola* to large geographical areas (Dixon et al., 2009; Sumabat et al., 2018). Therefore, genetic structure does not correlate with geographical origin of *C. cassiicola* isolates.

NLP-encoding genes in *C. cassiicola* evolved in a species-dependent manner and independently from each other

In a previous study on NLPs across Dothideomycetes, we have found evidences for the small size of the NLP family (three members) in *C. cassiicola* based on the genome of isolate CCP (Chapter 2). They were predicted to be effectors and each copy clustered in a different sub-clade among type I NLPs (Chapter 2). Herein, we chose *C. cassiicola* as a model to investigate the evolutionary history of NLPs in a species level.

We found that all three NLP effectors initially predicted in *C. cassiicola* CCP (Chapter 2) were retained by isolates of *C. cassiicola* from several host species and different geographic origins. Each NLP clade in the gene tree and haplogroup in network analysis were congruent with the species tree for *C. cassiicola*, suggesting each NLP gene evolved following the phylogenetic lineages of *C. cassiicola*. Clonal propagation, such as observed in *C. cassiicola* (Dixon et al., 2009), may favored for a rapid spread of favorable effector alleles within a species (Lo Presti et al., 2015). However, the only known species-specific effector from *C. cassiicola*, the cassiicolin, is an accessory effector that confers different virulence profiles to isolates (Déon et al., 2014; Lopez et al., 2018; Wu et al., 2018). Even isolates with no cassiicolin gene are able to cause disease (Déon et al., 2014). Thus, the maintenance of the three NLP effectors in *C. cassiicola* suggested that each of them play some important role in the pathogen-host interaction.

The large number of mutational steps that separate the NLP members in *C. cassiicola* and protein sequence variation suggest that each NLP effector gene has

followed an independent evolutionary path. The highly divergent members found in the NLP family from *C. cassiicola* supports the hypothesis for an ancient origin for the family (Oome and Van Den Ackerveken, 2014; Chapter 2). Additionally, the number of mutational steps separating haplotypes and high levels of nucleotide diversity within NLP members in *C. cassiicola* is congruent with the rapid evolution expected for fungal effectors (Lo Presti et al., 2015; Sánchez-Vallet et al., 2018).

The rare non-effector Cc_NLP1.4 possibly originated by a more recent duplication compared to other NLP effector members but have accumulated several mutations during its evolutionary path. Duplicated effector genes may be retained in response to selection imposed by the hosts and rapidly accumulate mutations (Fouché et al., 2018). During the last years, there has been a rapid spread of the target spot disease caused by *C. cassiicola* across soybean-growing regions in Brazil (Molina et al., 2019). It is plausible that rapid pathogen population growths have favored the emergence/maintenance of new genotypes in the pathogen population. However, our data do not allow us to confirm this hypothesis.

Selection upon NLP effector-encoding genes of *C. cassiicola*

We obtained evidence that NLP effector-encoding genes of *C. cassiicola* are under different selective constraints. Although the intraspecific selective neutrality tests (Tajima's D, Fu and Li's D* and F*) indicated that Cc_NLP1.1 is evolving under neutrality, the lower levels of nucleotide diversity and the distribution of polymorphisms along the Cc_NLP1.1 gene suggest this locus is under stronger selective constraints. We previously showed that NLPs that are homologous to Cc_NLP1.1 have the broadest distribution among Dothideomycetes (Chapter 2) and, in species with duplicated NLPs, Cc_NLP1.1 homologues seem to be more cytotoxic to plants (Pour et al., 2020; Chapter 2). Therefore, it seems likely that selection is operating in different ways to eliminate deleterious polymorphisms from Cc_NLP1.1 haplotypes in *C. cassiicola.*

The excess of mutations found in Cc_NLP1.2 indicates this locus is evolving under more relaxed selective constraints compared to others NLP effector genes in *C. cassiicola*. Negative values of Fu and Li's D* and F* together with the high number of low frequency haplotypes and missing intermediate haplotypes suggest for a recent expansion in *C. cassiicola*. Accumulation of mutations as a source genetic of variability exposes effector genes to mutations with both beneficial and deleterious effects, that

at some point may be removed from population by strong selection (Fouché et al., 2018). Rapid evolution leading to an excess of deleterious mutations could eventually lead to loss-of-function of Cc_NLP1.2 in *C. cassiicola* and would help explain the rare occurrence of homologues to Cc_NLP1.2 to other species, as we previously reported (Chapter 2).

Selection analyses indicate that both balancing and positive selection are acting upon Cc_NLP1.3. Significant positive values estimated by selective neutrality tests (Tajima's D, Fu and Li's D* and F*) indicates that balancing selection is acting to maintain high levels of genetic variation among Cc_NLP1.3 haplotypes. Effector proteins under positive selection are often considered to operate in the frontline of host-pathogen interactions (Lo Presti et al., 2015). Our findings that Cc_NLP1.3 effector displays significant signals of positive selection agree with previous studies that identified positive selection in few (but not all) NLP genes of *Botrytis* spp. (Staats et al., 2007) and *Phytophthora sojae* (Dong et al., 2012). Furthermore, the corresponded residue to N77, positively selected in Cc_NLP1.3 from *C. cassiicola,* was identified to be under positive selection in a few NLP effectors from *P. sojae* (Dong et al., 2012). The biological implication of such positively selected sites remains elusive.

We showed that, regardless of being highly divergent, NLP effectors from *C. cassiicola* have a central segment of amino acids that are conserved. Previous reports demonstrated that this segment represents the part of NLPs capable of inducing typical MAMP-triggered defense responses in plants (Böhm et al., 2014; Oome et al., 2014). It suggests that the three NLP effectors from *C. cassiicola* are potentially functional and, to some extent, purifying selection is acting upon NLP effector genes to maintain these key residues free of changing (Rody and Oliveira, 2018; Terauchi and Yoshida, 2010).

NLP effector genes are early expressed during the host colonization by *C. cassiicola*

Expression analyses of NLP effector genes during the first 18 hours of host colonization indicated that NLPs are part of the effector armory of *C. cassiicola*, contributing to its pathogenicity. Although protein sequences have suggested that all NLP effectors of *C. cassiicola* are potentially functional, expression profiles indicated that only one (Cc_NLP1.1) is significantly expressed during the colonization of soybean plants at the early hours. A recent study also identified the homologue of

Cc_NLP1.1 in the isolate CCP as one of the effectors differentially expressed during the colonization of rubber leaves by *C. cassiicola* (Lopez et al., 2018). Different levels of expression for NLP gene members have been described by several species and, among them, Cc_NLP1.1 homologues usually displayed higher expression levels compared to other NLP members (reviewed in chapter 2).

Although lower expression levels were found in Cc_NLP1.2, we cannot affirm whether it happened due to the high divergence of the gene CC_29_g13117 compared to others Cc_NLP1.2 homologues of *C. cassiicola* or due to relaxed selective constraints acting upon the Cc_NLP1.2 members. Nevertheless, the present results reinforce the hypothesis that accumulation of mutations increases the chance of emergence of deleterious mutation in effector genes, which in turn can lead to subfunctionalization or loss-of-function in effector genes (Fouché et al., 2018). In summary, the different levels of expression we uncovered among NLP effector genes suggested two non-mutually exclusive scenarios: 1) the NLP effector genes are expressed in different levels during host colonization; 2) not all NLP effector genes are essential for pathogenicity of *C. cassiicola* against soybean. Our data does not allow us to differentiate between these two scenarios.

In conclusion, the small size of NLP family in *C. cassiicola* makes the pathogen an interest model to investigate the evolutionary history of NLP effectors in a species framework. We showed that each NLP members evolved following the phylogenetic lineages of *C. cassiicola* and are under different selective constraints from each other. Although highly divergent, NLP effector genes maintained the functional part of the protein conserved, suggesting all NLP effectors are capable of inducing defense responses in plants. However, we uncovered different levels of expression among NLP effector genes, suggesting that Cc_NLP1.1 is significantly involved in the pathogenicity of *C. cassiicola* during the first hours of colonization. The work we present herein shed light on the evolution of NLP effector genes at a species levels and provided information on the gene genealogy of NLP genes in the ubiquitous plant pathogen *C. cassiicola*.

Acknowledgement

We thank Dr. Eveline Caixeta and other members of the Laboratório de Biotecnologia do Cafeeiro for the use of the Real-Time PCR System. This work was supported by The Minas Gerais State Foundation of Research Aid – FAPEMIG (grant number APQ-00150-17) and by The National Council of Scientific and Technological Development – CNPq (fellowship number PQ 302336/2019-2) to LOO. TCSD received student fellowships from the CAPES Foundation (PROEX – 0487 No. 1684083) from Mar/2017 to Jul/2018 and CNPq (GM/GD 142400/2018-1) from Aug/2018 to Apr/2021.

Conflict of interest

The authors have declared that no competing interests exist.

REFERENCES

Bae, Hanhong, Kim, M.S., Sicher, R.C., Bae, Hyeun-jong, Bailey, B.A., 2006. Necrosis- and Ethylene-inducing peptide from *Fusarium oxysporum* induces a complex cascade of transcripts associated with signal transduction and cell death in *Arabidopsis*. Plant Physiol. 141, 1056–1067.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S., 2009. MEME Suite: Tools for motif discovery and searching. Nucleic Acids Res. 37, 202–208.

Bandelt, H.J., Forster, P., Rohl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37–48.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477.

Böhm, H., Albert, I., Oome, S., Raaymakers, T.M., Van den Ackerveken, G., Nürnberger, T., 2014. A conserved peptide pattern from a widespread microbial virulence factor triggers pattern-induced immunity in *Arabidopsis*. PLoS Pathog. 10.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Breton, F., Sanier, C., D'Auzac, J., 2000. Role of cassiicolin, a host-selective toxin in pathogenicity *of C. cassiicola*, causal agent of leaf disease of *Hevea*. J. Nat. Rubber Res. 3, 115–128.

Buchfink, B., Xie, C., Huson, D.H., 2014. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60.

Cabral, A., Oome, S., Sander, N., Küfner, I., Nürnberger, T., Van Den Ackerveken, G., 2012. Nontoxic Nep1-like proteins of the downy mildew pathogen *Hyaloperonospora arabidopsidis*: Repression of necrosis-inducing activity by a surface-exposed region. Mol. Plant-Microbe Interact. 25, 697–708.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.

de Lamotte, F., Duviau, M.P., Sanier, C., Thai, R., Poncet, J., Bieysse, D., Breton, F., Pujade-Renaud, V., 2007. Purification and characterization of cassiicolin, the toxin produced by *Corynespora cassiicola*, causal agent of the leaf fall disease of rubber tree. J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 849, 357–362.

Déon, M., Bourré, Y., Gimenez, S., Berger, A., Bieysse, D., De Lamotte, F., Poncet, J., Roussel, V., Bonnot, F., Oliver, G., Franchel, J., Seguin, M., Leroy, T., Roeckel-Drevet, P., Pujade-Renaud, V., 2012a. Characterization of a cassiicolin-encoding gene from *Corynespora cassiicola*, pathogen of rubber tree (*Hevea brasiliensis*). Plant Sci. 185–186, 227–237.

Déon, M., Fumanal, B., Gimenez, S., Bieysse, D., Oliveira, R.R., Shuib, S.S., Breton, F., Elumalai, S., Vida, J.B., Seguin, M., Leroy, T., Roeckel-Drevet, P., Pujade-Renaud, V., 2014. Diversity of the cassiicolin gene in *Corynespora cassiicola* and relation with the pathogenicity in *Hevea brasiliensis*. Fungal Biol. 118, 32–47.

Déon, M., Scomparin, A., Tixier, A., Mattos, C.R.R., Leroy, T., Seguin, M., Roeckel-Drevet, P., Pujade-Renaud, V., 2012b. First characterization of endophytic *Corynespora cassiicola* isolates with variant cassiicolin genes recovered from

rubber trees in Brazil. Fungal Divers. 54, 87–99.

Dixon, L.J., Schlub, R.L., Pernezny, K., Datnoff, L.E., 2009. Host specialization and phylogenetic diversity of *Corynespora cassiicola*. Phytopathology 99, 1015–1027.

Dong, S., Kong, G., Qutob, D., Yu, X., Tang, J., Kang, J., Dai, T., Wang, H., Gijzen, M., Wang, Y., 2012. The NLP toxin family in *Phytophthora sojae* includes rapidly evolving groups that lack necrosis-inducing activity. Mol. Plant-Microbe Interact. 25, 896–909.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., Finn, R.D., 2019. The Pfam protein families database in 2019. Nucleic Acids Res. 47, D427–D432.

Emanuelsson, O., Brunak, S., von Heijne, G., Nielsen, H., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. 2, 953–971.

Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 1–14.

Fellbrich, G., Romanski, A., Varet, A., Blume, B., Brunner, F., Engelhardt, S., Felix, G., Kemmerling, B., Krzymowska, M., Nürnberger, T., 2002. NPP1, a *Phytophthora*-associated trigger of plant defense in parsley and *Arabidopsis*. Plant J. 32, 375–390.

Fouché, S., Plissonneau, C., Croll, D., 2018. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. Curr. Opin. Microbiol. 46, 34–42.

Franceschetti, M., Maqbool, A., Jiménez-Dalmaroni, M.J., Pennington, H.G., Kamoun, S., Banfield, M.J., 2017. Effectors of filamentous plant pathogens: Commonalities amid diversity. Microbiol. Mol. Biol. Rev. 81, 1–17.

Fu, Y.X., Li, W.H., 1993. Statistical tests of neutrality of mutations. Genetics 133, 693 LP – 709.

Gao, S., Zeng, R., Xu, L., Song, Z., Gao, P., Dai, F., 2020. Genome sequence and spore germination-associated transcriptome analysis of *Corynespora cassiicola* from cucumber. BMC Microbiol. 20, 1–20.

Gijzen, M., Nürnberger, T., 2006. Nep1-like proteins from plant pathogens: Recruitment and diversification of the NPP1 domain across taxa. Phytochemistry 67, 1800–1807.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by Maximum Likelihood. Syst. Biol. 52, 696–704.

Jones, D.A., Bertazzoni, S., Turo, C.J., Syme, R.A., Hane, J.K., 2018. Bioinformatic prediction of plant–pathogenicity effector proteins of fungi. Curr. Opin. Microbiol. 46, 43–49.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: Genome-scale protein function classification. Bioinformatics 30, 1236–1240.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E.L.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. 305, 567–580.

Lenarčič, T., Albert, I., Böhm, H., Hodnik, V., Pirc, K., Zavec, A.B., Podobnik, M., Pahovnik, D., Ema, Ž., Pruitt, R., Greimel, P., Yamaji-hasegawa, A., Kobayashi, T., 2017. Eudicot plant-specific sphingolipids determine host selectivity of microbial NLP cytolysins. Science. 358, 1431–1434.

Liu, L., Xu, L., Jia, Q., Pan, R., Oelmüller, R., Zhang, W., Wu, C., 2019. Arms race: diverse effector proteins with conserved motifs. Plant Signal. Behav. 14.

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., Kahmann, R., 2015. Fungal effectors and plant susceptibility. Annu. Rev. Plant Biol. 66, 513–545.

Looi, H.K., Toh, Y.F., Yew, S.M., Na, S.L., Tan, Y.C., Chong, P.S., Khoo, J.S., Yee, W.Y., Ng, K.P., Kuan, C.S., 2017. Genomic insight into pathogenicity of dematiaceous fungus *Corynespora cassiicola*. PeerJ 5, 1–28.

Lopez, D., Ribeiro, S., Label, P., Fumanal, B., Venisse, J.S., Kohler, A., Oliveira, R.R., Labutti, K., Lipzen, A., Lail, K., Bauer, D., Ohm, R.A., Barry, K.W., Spatafora, J., Grigoriev, I. V., Martin, F.M., Pujade-Renaud, V., 2018. Genome-wide analysis of *Corynespora cassiicola* leaf fall disease putative effectors. Front. Microbiol. 9, 1–21.

McDonald, J., Kreitman, H., 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351, 652–654.

Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., Gurevich, A., 2018. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics 34, i142–i150.

Molina, J.P.E., Paul, P.A., Amorim, L., da Silva, L.H.C.P., Siqueri, F. V., Borges, E.P., Campos, H.D., Venancio, W.S., Meyer, M.C., Martins, M.C., Balardin, R.S., Carlin, V.J., Grigolli, J.F.J., Belufi, L.M. d. R., Nunes Junior, J., Godoy, C. V., 2019. Effect of target spot on soybean yield and factors affecting this relationship. Plant Pathol. 68, 107–115.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.

Nylander, J.A.A., 2004. MrModeltest v2. Program distributed by the author.

Oome, S., Raaymakers, T.M., Cabral, A., Samwel, S., Böhm, H., Albert, I., Nürnberger, T., Van Den Ackerveken, G., 2014. Nep1-like proteins from three kingdoms of life act as a microbe-associated molecular pattern in *Arabidopsis*. Proc. Natl. Acad. Sci. U. S. A. 111, 16955–16960.

Oome, S., Van Den Ackerveken, G., 2014. Comparative and functional analysis of the widely occurring family of Nep1-like proteins. Mol. Plant-Microbe Interact. 27, 1081–1094.

Ottmann, C., Luberacki, B., Küfner, I., Koch, W., Brunner, F., Weyand, M., Mattinen, L., Pirhonen, M., Anderluh, G., Seitz, H.U., Nürnberger, T., Oecking, C., 2009. A common toxin fold mediates microbial attack and plant defense. Proc. Natl. Acad. Sci. U. S. A. 106, 10359–10364.

Petersen, T.N., Brunak, S., Von Heijne, G., Nielsen, H., 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.

Pour, F.N., Cobos, R., Coque, J.J.R., Serôdio, J., Alves, A., Félix, C., Ferreira, V., Esteves, A.C., Duarte, A.S., 2020. Toxicity of recombinant Necrosis and Ethylene-inducing proteins (NLPs) from *Neofusicoccum parvum*. Toxins. 12, 1–17.

Qutob, D., Kemmerling, B., Brunner, F., Küfner, I., Engelhardt, S., Gust, A.A., Luberacki, B., Seitz, H.U., Stahl, D., Rauhut, T., Glawischnig, E., Schween, G., Lacombe, B., Watanabe, N., Lam, E., Schlichting, R., Scheel, D., Nau, K., Dodt, G., Hubert, D., Gijzen, M., Nürnberger, T., 2006. Phytotoxicity and innate immune responses induced by Nep1-like proteins. Plant Cell 18, 3721–3744.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67, 901–904.

Ramírez-Soriano, A., Ramos-Onsins, S.E., Rozas, J., Calafell, F., Navarro, A., 2008. Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. Genetics 179, 555–567.

Rody, H.V.S., Oliveira, L.O., 2018. Evolutionary history of the cobalamin-independent methionine synthase gene family across the land plants. Mol. Phylogenet. Evol. 120, 33–42.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., Sánchez-Gracia, A., 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. 34, 3299–3302.

Sánchez-Vallet, A., Fouché, S., Fudal, I., Hartmann, F.E., Soyer, J.L., Tellier, A., Croll, D., 2018. The genome biology of effector gene evolution in filamentous plant pathogens. Annu. Rev. Phytopathol. 56.

Seidl, M.F., Van den Ackerveken, G., 2019. Activity and phylogenetics of the broadly occurring family of microbial Nep1-Like Proteins. Annu. Rev. Phytopathol. 57, 367–386.

Shrestha, S.K., Lamour, K., Young-Kelly, H., 2017. Genome sequences and SNP analyses of *Corynespora cassiicola* from cotton and soybean in the southeastern United States reveal limited diversity. PLoS One 12, 6–14.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E.M., 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.

Sonah, H., Deshmukh, R.K., Bélanger, R.R., 2016. Computational prediction of effector proteins in fungi: Opportunities and challenges. Front. Plant Sci. 7, 1–14.

Staats, M., van Baarlen, P., Schouten, A., van Kan, J.A.L., Bakker, F.T., 2007. Positive selection in phytotoxic protein-encoding genes of *Botrytis* species. Fungal Genet. Biol. 44, 52–63.

Stanke, M., Morgenstern, B., 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33, 465–467.

Sumabat, L.G., Kemerait, R.C., Brewer, M.T., 2018. Phylogenetic diversity and host specialization of *Corynespora cassiicola* responsible for emerging target spot disease of cotton and other crops in the southeastern United States. Phytopathology 108, 892–901.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Terauchi, R., Yoshida, K., 2010. Towards population genomics of effector-effector target interactions. New Phytol. 187, 929–939.

Wu, J., Xie, X., Shi, Y., Chai, A., Wang, Q., Li, B., 2018. Variation of cassiicolin genes among Chinese isolates of *Corynespora cassiicola*. J. Microbiol. 56, 634–647.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.

Supplementary material

Evolutionary history and expression analysis of the Necrosis- and Ethylene-inducing peptide 1-like protein (NLP) gene family in *Corynespora cassiicola*

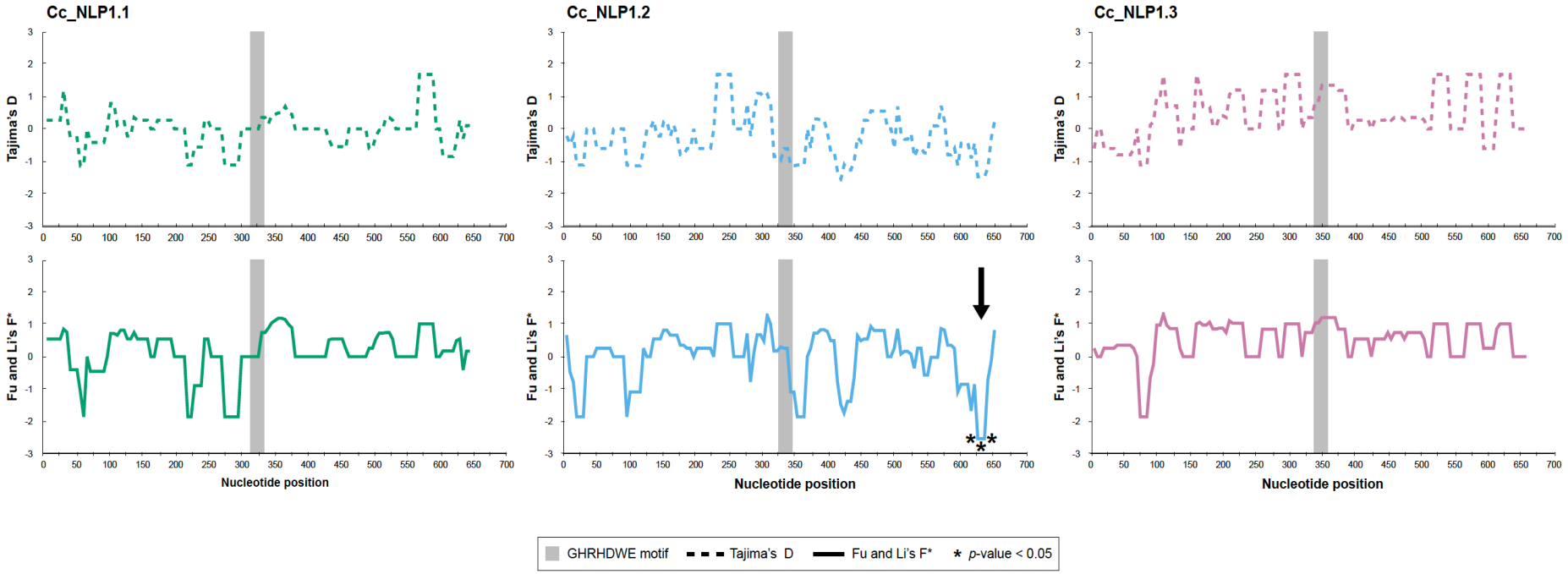


Figure S1. Sliding window neutrality tests of three predicted Necrosis- and Ethylene-inducing peptide 1-like protein (NLP)-encoding genes (CDS) of *Corynespora cassiicola*. The x axis shows the midpoints of contiguous windows (25 bases, with steps of 5 bases). The grey vertical bar indicates the relative position of the GHRHDWE motif. The arrow indicates sites with significant departures from zero (*p*-value < 0.05).

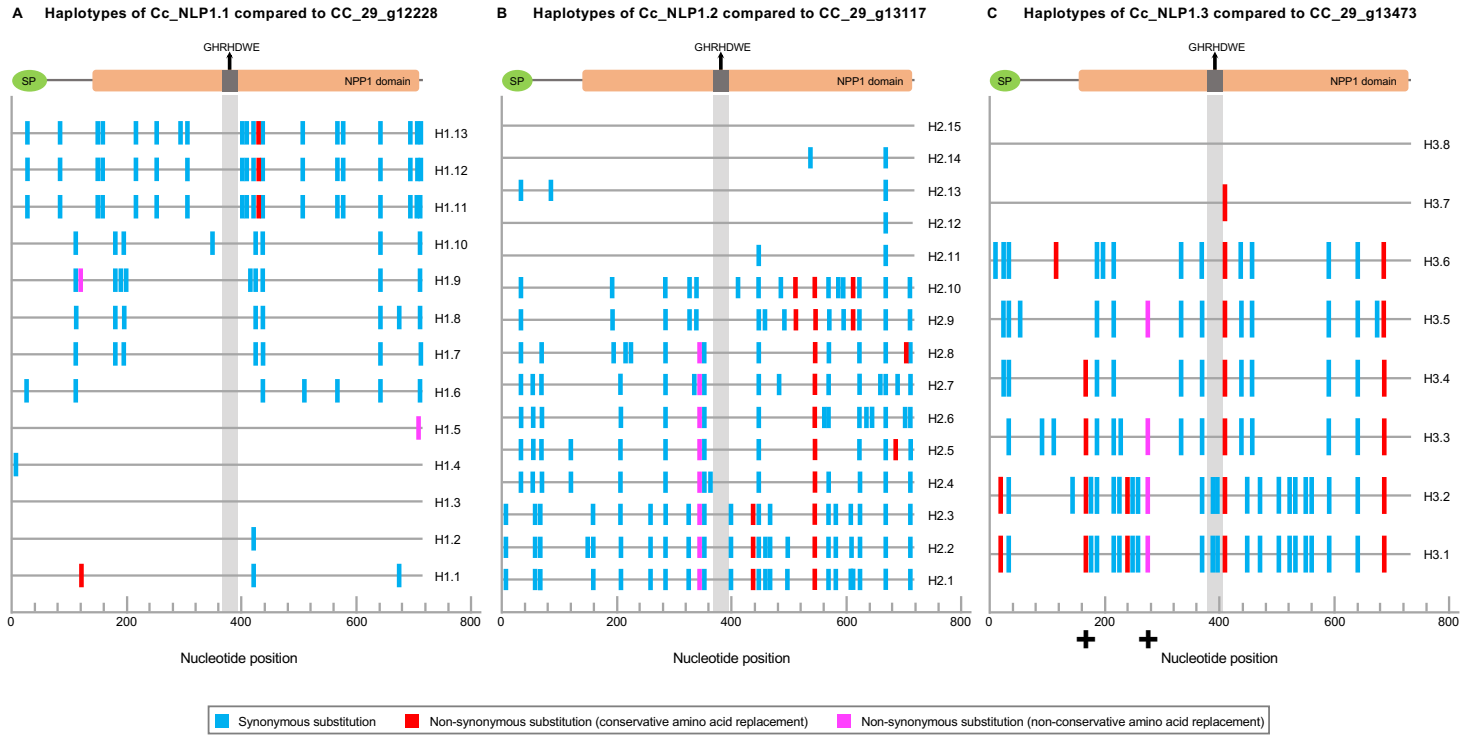


Figure S2. Placement of nucleotide substitutions along the haplotypes of each of the three NLP effector-encoding genes (Cc_NLP1.1, Cc_NLP1.2, and Cc_NLP1.3) (CDS) in *Corynespora cassiicola.* The NLP effector-encoding genes from isolate CC_29 were used as reference for haplotype sequence alignment, according to the haplogroups in Figure 3. Upper protein representations indicate portions of the CDS that encode for the signal peptide (SP), the NPP1 domain, and the conserved GHRHDWE motif in the NLP effectors of isolate CC_29. **A.** Cc_NLP1.1 compared to CC_29_g12228 (haplotype 1.3). **B**. Cc_NLP1.2 compared to CC_29_g13117 (haplotype 2.15). **C.** Cc_NLP1.3 compared to CC_29_g13473 (haplotype 3.8). Plus (+) signs indicate substitutions that resulted in amino acids replacements predicted to be positively selected, according to *codeml*/PAML analysis. Types of nucleotide substitutions as indicated.

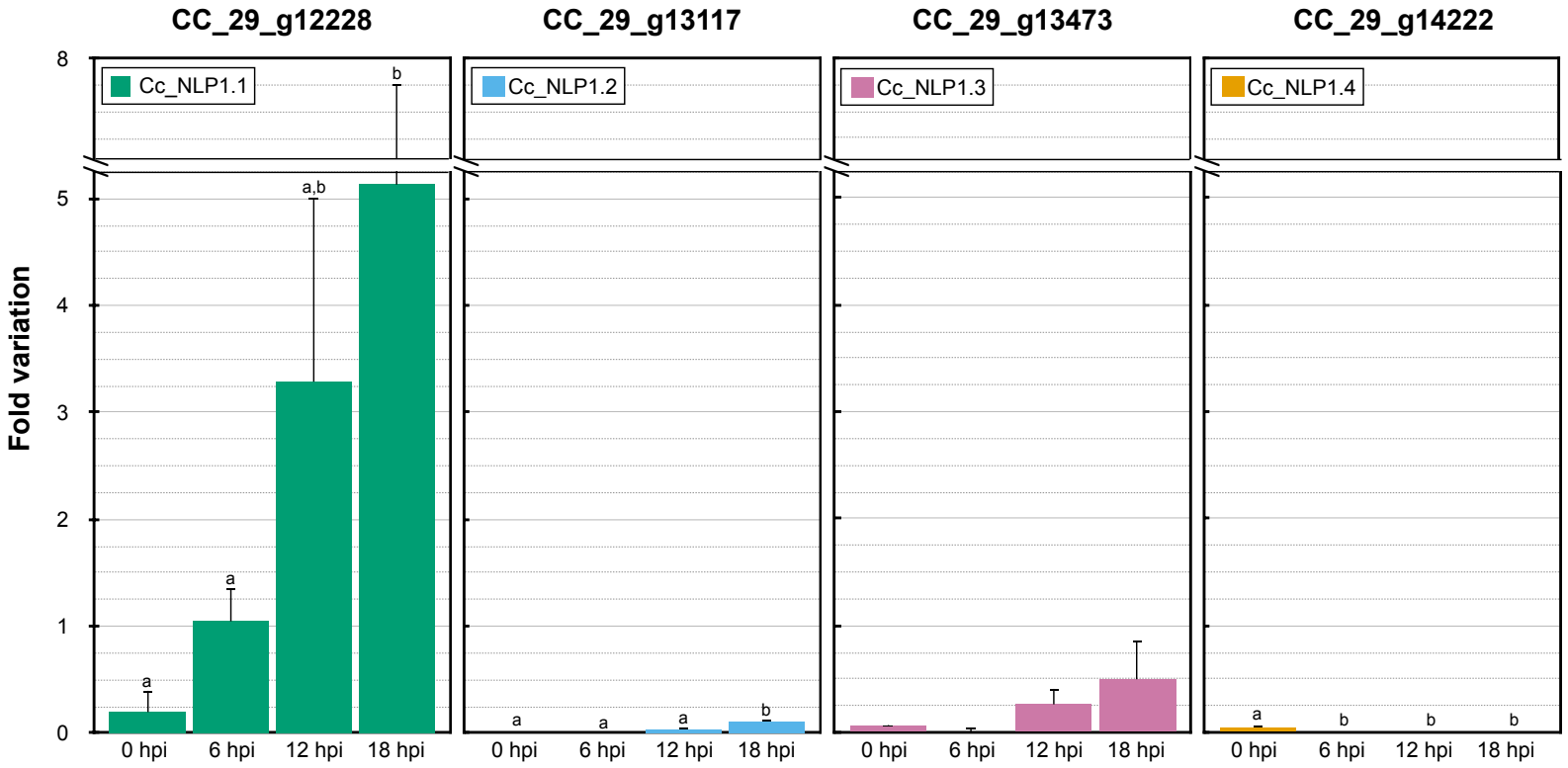


Figure S3. Expression patterns of the Necrosis- and Ethylene-inducing peptide 1-like (NLP) genes of *Corynespora cassiicola* (isolate CC_29) after inoculation in soybean plants. Gene expression was inferred by RT-qPCR analyses. The time (x axis) is given in hours post inoculation (hpi). The fold variation of gene expression of each NLP gene (y axis) was calculated using the $2^{-\Delta Ct}$ method, the values were normalized to an endogenous control gene (β-tubulin). Letters show significant differences (p-value < 0.05) after a one-way ANOVA followed by Tukey's test for each NLP gene individually. Error bars represent the standard deviation of three biological replicates.

Table S1. Detailed information on all isolates of *Corynespora cassiicola* used in this study.

Isolate	BioSample	Host Species	Geographic Origin	Reference	Genome Size (Mb)	# Genes	# NLP Genes	# NLP Effectors	NLP Gene ID	Haplogroup	Haplotype
777AA	SAMN08330686	*Glycine max*	Brazil	Lopez et al. (2018)	41.70	14,873	4	3	777AA_g10179	Cc_NLP1.1	1.2
									777AA_g880	Cc_NLP1.2	2.11
									777AA_g10197	Cc_NLP1.3	3.8
									777AA_g14294	Cc_NLP1.4	4.1
ATI11	SAMN08330687	*Cucumis sativus*	Brazil	Lopez et al. (2018)	41.13	14,497	3	3	ATI11_g10166	Cc_NLP1.1	1.3
									ATI11_g1730	Cc_NLP1.2	2.15
									ATI11_g11360	Cc_NLP1.3	3.8
C7_L008	SAMEA103891068	-	-	-	42.50	14,569	3	3	C7_L008_g6805	Cc_NLP1.1	1.4
									C7_L008_g10130	Cc_NLP1.2	2.11
									C7_L008_g6820	Cc_NLP1.3	3.7
CAL-4	SAMN12086008	*Gossypium hirsutum*	United States	-	44.90	19,240	3	3	CAL-4_g12966	Cc_NLP1.1	1.2
									CAL-4_g6936	Cc_NLP1.2	2.11
									CAL-4_g13172	Cc_NLP1.3	3.8
CBS129.25	SAMN08330688	*Cucumis sativus*	Brazil	Lopez et al. (2018)	40.79	14,403	3	3	CBS12925_g6790	Cc_NLP1.1	1.2
									CBS12925_g3518	Cc_NLP1.2	2.15
									CBS12925_g6808	Cc_NLP1.3	3.8
CC_29	SAMN18718492	*Glycine max*	Brazil	Chapter 1	44.79	18,487	4	3	CC_29_g12228	Cc_NLP1.1	1.3
									CC_29_g13117	Cc_NLP1.2	2.15
									CC_29_g13473	Cc_NLP1.3	3.8
									CC_29_g14222	Cc_NLP1.4	4.3
CCAM1	SAMN08330689	*Hevea brasiliensis*	Cameroon	Lopez et al. (2018)	42.33	14,634	3	3	CCAM1_g9108	Cc_NLP1.1	1.7
									CCAM1_g5347	Cc_NLP1.2	2.4
									CCAM1_g13882	Cc_NLP1.3	3.4
CCAM2	SAMN08330690	*Hevea brasiliensis*	Cameroon	Lopez et al. (2018)	41.89	14,785	3	3	CCAM2_g9764	Cc_NLP1.1	1.11
									CCAM2_g3518	Cc_NLP1.2	2.6
									CCAM2_g10878	Cc_NLP1.3	3.5
CCAM3	SAMN08330691	*Hevea brasiliensis*	Cameroon	Lopez et al. (2018)	42.72	15,022	3	3	CCAM3_g102	Cc_NLP1.1	1.11
									CCAM3_g10005	Cc_NLP1.2	2.6
									CCAM3_g12335	Cc_NLP1.3	3.5
CCAM4	SAMN08330692	*Hevea brasiliensis*	Cameroon	Lopez et al. (2018)	42.32	14,545	3	3	CCAM4_g6615	Cc_NLP1.1	1.7
									CCAM4_g3778	Cc_NLP1.2	2.4
									CCAM4_g9511	Cc_NLP1.3	3.4
CCI13	SAMN08330693	*Hevea brasiliensis*	Côte d'Ivoire	Lopez et al. (2018)	41.94	14,847	3	3	CCI13_g7758	Cc_NLP1.1	1.3
									CCI13_g2701	Cc_NLP1.2	2.11
									CCI13_g12991	Cc_NLP1.3	3.7
CCI6	SAMN08330694	*Hevea brasiliensis*	Côte d'Ivoire	Lopez et al. (2018)	41.99	14,905	3	3	CCI6_g6508	Cc_NLP1.1	1.3
									CCI6_g12011	Cc_NLP1.2	2.11
									CCI6_g10427	Cc_NLP1.3	3.7
CCP	SAMN05660679	*Hevea brasiliensis*	Philippines	Lopez et al. (2018)	44.85	17,166	3	3	CCP_601720	Cc_NLP1.1	1.11
									CCP_537044	Cc_NLP1.2	2.6
									CCP_497672	Cc_NLP1.3	3.5
CGAB1	SAMN08330695	*Hevea brasiliensis*	Gabon	Lopez et al. (2018)	41.20	14,635	3	3	CGAB1_g7423	Cc_NLP1.1	1.3
									CGAB1_g5586	Cc_NLP1.2	2.14
									CGAB1_g8330	Cc_NLP1.3	3.7
CGAB2	SAMN08330696	*Hevea brasiliensis*	Gabon	Lopez et al. (2018)	41.22	14,638	3	3	CGAB2_g869	Cc_NLP1.1	1.3
									CGAB2_g1004	Cc_NLP1.2	2.14
									CGAB2_g3883	Cc_NLP1.3	3.7

Table S1. Continued.

Isolate	BioSample	Host Species	Geographic Origin	Reference	Genome Size (Mb)	# Genes	# NLP Genes	# NLP Effectors	NLP Gene ID	Haplogroup	Haplotype
CIND3	SAMN08330697	*Hevea brasiliensis*	India	Lopez et al. (2018)	41.67	14,733	3	3	CIND3_g13589	Cc_NLP1.1	1.3
									CIND3_g13380	Cc_NLP1.2	2.11
									CIND3_g14081	Cc_NLP1.3	3.7
CLN16	SAMN08330698	*Hevea brasiliensis*	Malaysia	Lopez et al. (2018)	41.71	14,803	3	3	CLN16_g9268	Cc_NLP1.1	1.3
									CLN16_g9820	Cc_NLP1.2	2.11
									CLN16_g10739	Cc_NLP1.3	3.7
CSB16	SAMN08330699	*Hevea brasiliensis*	Malaysia	Lopez et al. (2018)	40.65	14,489	3	3	CSB16_g4822	Cc_NLP1.1	1.12
									CSB16_g11473	Cc_NLP1.2	2.9
									CSB16_g4841	Cc_NLP1.3	3.1
CSRI1	SAMN08330700	*Hevea brasiliensis*	Sri Lanka	Lopez et al. (2018)	40.61	14,327	3	3	CSRI1_g7251	Cc_NLP1.1	1.12
									CSRI1_g13432	Cc_NLP1.2	2.9
									CSRI1_g7232	Cc_NLP1.3	3.1
CSRI2	SAMN08330701	*Hevea brasiliensis*	Sri Lanka	Lopez et al. (2018)	40.77	14,391	3	3	CSRI2_g9601	Cc_NLP1.1	1.12
									CSRI2_g8039	Cc_NLP1.2	2.9
									CSRI2_g9582	Cc_NLP1.3	3.1
CSRI5	SAMN08330702	*Hevea brasiliensis*	Sri Lanka	Lopez et al. (2018)	41.67	14,736	3	3	CSRI5_g4338	Cc_NLP1.1	1.3
									CSRI5_g10755	Cc_NLP1.2	2.11
									CSRI5_g9398	Cc_NLP1.3	3.7
CTHA1	SAMN08330703	*Hevea brasiliensis*	Thailand	Lopez et al. (2018)	41.28	14,432	3	3	CTHA1_g5888	Cc_NLP1.1	1.9
									CTHA1_g12212	Cc_NLP1.2	2.8
									CTHA1_g1711	Cc_NLP1.3	3.6
CTHA2	SAMN08330704	*Hevea brasiliensis*	Thailand	Lopez et al. (2018)	40.82	14,424	3	3	CTHA2_g8146	Cc_NLP1.1	1.12
									CTHA2_g8485	Cc_NLP1.2	2.9
									CTHA2_g8165	Cc_NLP1.3	3.1
CTHA3	SAMN08330705	*Hevea brasiliensis*	Thailand	Lopez et al. (2018)	40.37	14,616	3	3	CTHA3_g11953	Cc_NLP1.1	1.9
									CTHA3_g5642	Cc_NLP1.2	2.8
									CTHA3_g4093	Cc_NLP1.3	3.6
CTHA4	SAMN08330706	*Hevea brasiliensis*	Thailand	Lopez et al. (2018)	40.71	14,410	3	3	CTHA4_g13560	Cc_NLP1.1	1.9
									CTHA4_g2821	Cc_NLP1.2	2.8
									CTHA4_g6803	Cc_NLP1.3	3.6
CTHA5	SAMN08330707	*Hevea brasiliensis*	Thailand	Lopez et al. (2018)	40.86	14,461	3	3	CTHA5_g13106	Cc_NLP1.1	1.5
									CTHA5_g688	Cc_NLP1.2	2.11
									CTHA5_g13694	Cc_NLP1.3	3.7
CTHA6	SAMN08330708	*Hevea brasiliensis*	Thailand	Lopez et al. (2018)	40.94	14,444	3	3	CTHA6_g4937	Cc_NLP1.1	1.3
									CTHA6_g646	Cc_NLP1.2	2.11
									CTHA6_g11768	Cc_NLP1.3	3.7
CVa5	SAMN12086010	*Gossypium hirsutum*	United States	-	44.52	15,348	3	3	CVa5_g12316	Cc_NLP1.1	1.2
									CVa5_g12873	Cc_NLP1.2	2.11
									CVa5_g13757	Cc_NLP1.3	3.8
E139	SAMN08330709	*Hevea brasiliensis*	Brazil	Lopez et al. (2018)	42.82	14,648	3	3	E139_g2766	Cc_NLP1.1	1.8
									E139_g2931	Cc_NLP1.2	2.7
									E139_g11927	Cc_NLP1.3	3.4
E55	SAMN08330710	*Hevea brasiliensis*	Brazil	Lopez et al. (2018)	40.53	14,352	3	3	E55_g11737	Cc_NLP1.1	1.6
									E55_g6497	Cc_NLP1.2	2.2
									E55_g10462	Cc_NLP1.3	3.3

Table S1. Continued.

Isolate	BioSample	Host Species	Geographic Origin	Reference	Genome Size (Mb)	# Genes	# NLP Genes	# NLP Effectors	NLP Gene ID	Haplogroup	Haplotype
E79	SAMN08330711	*Hevea brasiliensis*	Brazil	Lopez et al. (2018)	42.47	14,373	3	3	E79_g12376	Cc_NLP1.1	1.10
									E79_g1196	Cc_NLP1.2	2.5
									E79_g5219	Cc_NLP1.3	3.4
EDIG	SAMN08330712	*Cucumis sativus*	Brazil	Lopez et al. (2018)	41.24	14,743	3	3	EDIG_g11647	Cc_NLP1.1	1.3
									EDIG_g11366	Cc_NLP1.2	2.15
									EDIG_g6937	Cc_NLP1.3	3.8
GSO2	SAMN08330713	*Vernonia cinerea*	Brazil	Lopez et al. (2018)	41.80	15,072	3	3	GSO2_g7213	Cc_NLP1.1	1.3
									GSO2_g1195	Cc_NLP1.2	2.12
									GSO2_g529	Cc_NLP1.3	3.7
IA	SAMN08330714	*Cucumis sativus*	Brazil	Lopez et al. (2018)	41.58	14,793	3	3	IA_g12863	Cc_NLP1.1	1.2
									IA_g8583	Cc_NLP1.2	2.15
									IA_g14565	Cc_NLP1.3	3.8
India_Hevea	SAMN10434155	*Hevea brasiliensis*	India	-	42.89	14,404	3	3	India_Hevea_g13512	Cc_NLP1.1	1.1
									India_Hevea_g4639	Cc_NLP1.2	2.13
									India_Hevea_g13452	Cc_NLP1.3	3.8
JQ	SAMN08330715	*Cucumis sativus*	Brazil	Lopez et al. (2018)	41.55	14,815	3	3	JQ_g1090	Cc_NLP1.1	1.2
									JQ_g6266	Cc_NLP1.2	2.15
									JQ_g14090	Cc_NLP1.3	3.8
LPO7	SAMN08330716	*Piper hispidinervum*	Brazil	Lopez et al. (2018)	42.23	14,535	3	3	LPO7_g9249	Cc_NLP1.1	1.6
									LPO7_g5694	Cc_NLP1.2	2.1
									LPO7_g9266	Cc_NLP1.3	3.3
PB	SAMN08330717	*Cucumis sativus*	Brazil	Lopez et al. (2018)	41.56	14,801	3	3	PB_g7351	Cc_NLP1.1	1.2
									PB_g3886	Cc_NLP1.2	2.15
									PB_g2314	Cc_NLP1.3	3.8
RUD	SAMN08330718	*Glycine max*	Brazil	Lopez et al. (2018)	41.63	14,858	4	3	RUD_g9099	Cc_NLP1.1	1.3
									RUD_g11509	Cc_NLP1.2	2.15
									RUD_g10304	Cc_NLP1.3	3.8
									RUD_g11851	Cc_NLP1.4	4.2
SS1	SAMN08330719	*Hevea brasiliensis*	Malaysia	Lopez et al. (2018)	40.73	14,378	3	3	SS1_g6084	Cc_NLP1.1	1.12
									SS1_g6638	Cc_NLP1.2	2.9
									SS1_g6065	Cc_NLP1.3	3.1
TCl3	SAMN12086019	*Solanum lycopersicum*	United States	-	45.34	14,987	3	3	TCl3_g7271	Cc_NLP1.1	1.6
									TCl3_g7342	Cc_NLP1.2	2.3
									TCl3_g12718	Cc_NLP1.3	3.3
TSB1	SAMN08330720	*Hevea brasiliensis*	Malaysia	Lopez et al. (2018)	40.89	14,435	3	3	TSB1_g6254	Cc_NLP1.1	1.12
									TSB1_g3780	Cc_NLP1.2	2.9
									TSB1_g6273	Cc_NLP1.3	3.1
TScotton1	SAMN06706475	*Gossypium hirsutum*	United States	Shrestha et al. (2018)	42.10	14,988	3	3	TScotton1_g11067	Cc_NLP1.1	1.2
									TScotton1_g11728	Cc_NLP1.2	2.11
									TScotton1_g11410	Cc_NLP1.3	3.8
UM591	SAMN02981578	*Homo sapiens*	Malaysia	Looi et al. (2017)	41.35	14,695	3	3	UM591_g2079	Cc_NLP1.1	1.13
									UM591_g4896	Cc_NLP1.2	2.10
									UM591_g2097	Cc_NLP1.3	3.2

Table S2. Primers designed for *C. cassiicola* isolate CC_29 for NLP gene expression analyses by RT-qPCR.

Gene Name	Gene ID	Forward sequence (5' – 3')	Reverse sequence (5' – 3')
Cc_NLP1.1	CC_29_g12228	CCTGTCGAATCCTCCCTTGA	AACGACCGATAGCACCACTC
Cc_NLP1.2	CC_29_g13117	GAAGTTTGCGCCCTACATGC	GAGTCCACCGCTCGTGTTTC
Cc_NLP1.3	CC_29_g13473	GGGTTCCCTCAGACTGTTCC	ATGGCACGCAACCATTGAAG
Cc_NLP1.4	CC_29_g14222	CCATTGGCAGGGCTATGAAGA	CCAGCATCATCCTTAGCGGT
β-tubulin	CC_29_g4872	CAGACCGGTCAATGCGGTAA	GCCATTGTAGACGCCGGATC

Table S3. Summary statistics on *de novo* genome assemblies of *Corynespora cassiicola* isolates India_Hevea and TScotton1.

Genomic statistics	India_Hevea	TScotton1
Assembly		
Total assembly size (Mbp)	42.89	42.11
# of scaffolds (>= 500 bp)	1,059	2,188
# of scaffolds (>= 10,000 bp)	443	806
# of scaffolds (>= 50,000 bp)	246	282
Largest scaffold (bp)	827,165	251,612
GC (%)	50.85	52.56
N_{50} (bp)	149,160	62,602
Total reads	21,333,739	74,123,165
Mapped reads (%)	99.80	99.54
Average coverage depth (X)	74	263
Gene completeness		
Complete BUSCOs (%)	99.6	98.9
Complete and single-copy BUSCOs (%)	99.5	98.8

GENERAL CONCLUSIONS

GENERAL CONCLUSIONS

The two species *Corynespora cassiicola* isolate CC_29 and *Corynespora olivacea* isolate CBS 114450 differed greatly in genome sizes and effector content. The large difference is mostly due to the distant phylogenetic relationships between the two species. *Corynespora* is not a monophyletic genus and, taken together with previous reports, we suggested the classification of *C. olivacea* in the genus *Corynespora* should be revised.

The phylogenies of each NLP subfamilies followed closely the phylogenomic relationships of the genera in Dothideomycetes. Nevertheless, the imbalanced phylogeny of the NLP superfamily across the Dothideomycetes revealed that subfamilies underwent independent evolutionary paths, each of which showing different signatures of ancient gene duplications and biased successive gene losses that may be associated with changes in cytotoxic activity.

Our investigation on the evolutionary history of NLP effectors in a species framework shed further light in the activity of the NLP subfamilies. The small size of NLP family was maintained in isolates of *C. cassiicola,* where each NLP member evolved following the phylogenetic lineages of *C. cassiicola*. Stronger selective constraints and higher levels of gene expression found in NLP1.1 homologues in *C. cassiicola* reinforce the hypothesis of functionality for this subfamily across other taxa.